BUSINESS INTELLIGENCE PROGRAM

# Genomics: Technology Update 2002

October 2002

Andrew Broderick

**Insight**

D02-2401

SRI CONSULTING BUSINESS INTELLIGENCE

An SRI International Business Partner

© 2002 by SRI Consulting Business Intelligence. All rights reserved.

# GENOMICS: TECHNOLOGY UPDATE 2002

D02-2401

## CONTENTS


CONTINUING PROGRESS IN GENOMICS TECHNOLOGY ..... 1

SEQUENCE GENOMICS ..... 3

FUNCTIONAL GENOMICS ..... 4
Model Genetic Systems ..... 5
Gene-Expression Analysis ..... 6
Pharmacogenomics ..... 8
Proteomics ..... 9
Chemogenomics ..... 10
Toxicogenomics ..... 11
Epigenomics ..... 11
Metabolomics ..... 12
Systems Biology ..... 12

BIOINFORMATICS ..... 13

OUTLOOK ..... 14


### Tables


Levels of Detail in Functional Genomics ..... 5
Standards for Gene-Expression Data ..... 8


### Figure


Building Blocks of Genomics Technology ..... 2

# CONTINUING PROGRESS IN GENOMICS TECHNOLOGY

Genomics research has ushered in a new era of large-scale, automated, high-throughput science, which in turn has affected instrumentation and processes, ranging from sample preparation to data analysis. This work underscores the increasingly symbiotic relationship between technology development and scientific research: Advances in biology drive technology development as much as advances in technology drive biology. But implementing the opportunities that grow out of scientific research and discovery will be a challenge for companies, requiring them to unify many disciplines, scientific methods, organizations, emerging technologies, and levels of information across the research environment.

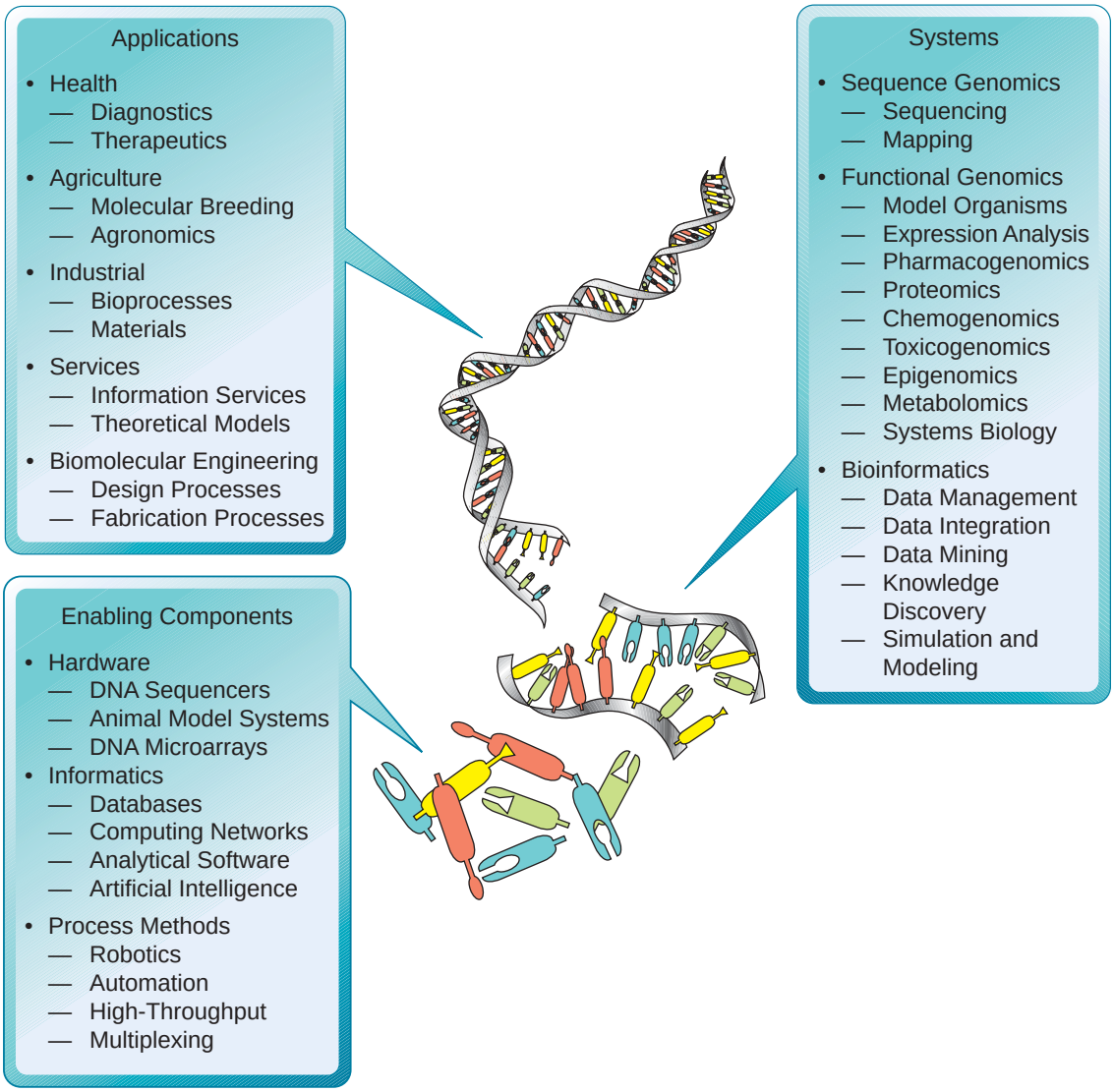
The Human Genome Project (HGP) is remarkable for its technical and scientific accomplishments. Yet despite new knowledge of the human genetic-sequence code, scientists still lack a clear understanding of the functions of most genes. Much of the scientific effort in the coming decades will aim to elucidate these functions, and the emphasis in research has already begun shifting away from sequence genomics toward functional genomics. Although functional genomics requires various scientific methods to elucidate gene function and the application of industrialization processes across the different disciplines will be common, advanced computational methods for the management and integration of data that will be generated by the variety of experimental techniques employed will be particularly critical.

Life-sciences researchers are working toward an integrative or systems-biology approach. To reach this goal, they need not only to build on technological developments in science but also to pursue interdisciplinary approaches in research. Standardization will likely be the key to unifying disparate areas of technical expertise into an integrated system for common understanding, and bioinformatics—the use of computers to analyze information emerging from the sequencing and mapping of genes and study of functional genomics—will be the key to building an enabling infrastructure. If researchers gain the ability to tie into a common knowledge infrastructure, the impact on science and medicine will be huge.

This study reviews current technologies and related developments in genomics research. Earlier studies of genomics technologies include:

- D01-2339, *Genomics: Technology Update 2001*
- D01-2315, *The Postgenomics Era: Proteomics and Structural Genomics*
- D00-2300, *DNA Chips: Technology Update*
- D00-2284, *Genomics: A Technology Update*
- D00-2254, *Bioinformatics: Emerging Role in Life-Sciences Research*
- D99-2215, *Bioinformatics: Making the Transition from Data to Knowledge*
- D98-2163, *Pharmacogenomics: Redefining the Pharmaceutical Industry.*

**Figure 1**
**BUILDING BLOCKS OF GENOMICS TECHNOLOGY**




Source: SRI Consulting Business Intelligence

## SEQUENCE GENOMICS

Though capillary-based instruments are today's state of the art in DNA sequencing, incremental improvements are under way to provide higher sample capacity, longer sequence-read lengths, and faster single-run times. The ability to reduce sample sizes will remove some of the costs of reagents and significantly reduce the time, effort, and labor costs of preparing, sequencing, and separating samples. Next-generation electrophoresis-based sequencers will most likely be miniaturized microchip-based systems that surpass current technology in speed, resolution, throughput, and detection sensitivity.

Recent advances in molecular biology have spurred the development of new methods of labeling and detecting DNA and DNA fragments. Traditionally, radioisotopes have served as labels for DNA. More recently, fluorescent, chemiluminescent, and bioactive reporter groups have found use. In the future, direct detection and analysis of single molecules using atomic-force microscopy or nanopore-sequencing methods will eliminate the need for labeling and theoretically provide a more sensitive and highly specific detection capability. If researchers can develop successful direct methods of DNA analysis in a format that is rapid, accurate, and cost-effective, they believe that the technology can eventually serve individual patients in clinical applications.

Many companies are working on ultra-high-throughput, miniaturized DNA-sequencing technology. U.S. Genomics (Woburn, Massachusetts) is developing technologies for direct linear analysis that can read long sequences of DNA, and Shimadzu Biotech (Pasadena, California) is working with microfabrication company GenoMEMS (Woburn, Massachusetts) on a new high-throughput DNA sequencer that it reports will be ten times faster and 90% cheaper than conventional systems. Solexa (Cambridge, England) is using "single-molecule arrays" that contain immobilized samples of single-strand DNA for sequence-variation analysis. Bioforce Nanosciences (Ames, Iowa) is developing nanoscale biomolecular-analysis technologies in arrays of less than 1 millimeter by using its proprietary NanoArrayer technology to deposit biomolecules onto its NanoArrays.

Curagen's (New Haven, Connecticut) subsidiary 454 Corporation is developing a unified platform of sequencing, profiling, and genotyping technologies to conduct whole-genome, massively parallel analysis, whereas Hyseq (Sunnyvale, California) has spun off Callida Genomics to develop and commercialize advanced DNA-sequence and -analysis tools using its proprietary sequencing-by-hybridization (SBH) technology. A subsidiary of Callida, N-Mer, has formed as part of a patent-litigation settlement with Affymetrix (Santa Clara, California): The company will have access to both Hyseq's SBH and Affymetrix's GeneChip technologies in developing a universal, high-speed DNA-sequencing chip. Nanosphere's (Northbrook, Illinois) system of gold nanoparticle probes uses electrical signals to detect the presence of target DNA strands; the system reportedly has the potential to be 10 times more sensitive and 100 000 times more selective than alternative DNA-detection technologies.

Although nanopore-based detection technology is unlikely to reach the market any time soon, nanoscale approaches to single-molecule analysis promise to provide an extremely rapid, label-free method of sequencing. Though some companies have been able to pass a DNA strand several thousand bases long through a nanopore, they have found that the DNA moves through the pore too quickly to allow direct reading of each base. Agilent Technologies (Palo Alto, California), which has teamed up with Harvard University's Daniel Branton to develop his group's nanopore-sequencing technology, plans to commercialize label-free nanopore-detection methods.

## FUNCTIONAL GENOMICS

Upon completion of the draft of the human genome, scientists still did not know the exact number of genes in the genome or the exact functions of genes. Because the study and determination of gene function typically use model organisms for comparative analyses and microarrays for gene-expression analyses, the comparative study of genomic information across species, individuals, and tissue samples will preoccupy researchers for years to come. Complementary scientific methods and disciplines are also emerging as research attention increasingly shifts toward finding answers to currently unknown scientific matters.

As the research focus shifts from problems of gene structure and genome organization to determination of the ways in which component parts function collectively, scientists will need to be able to study organisms in a high-throughput manner and from a global or systems perspective. Ultimately, researchers hope to use system-level information to reconstruct complex molecular networks. The ability to do so will enable them to develop models of these biological or cellular networks, which they can then use to generate hypotheses about disease processes that might ultimately lead to market breakthroughs in new therapeutics, diagnostics, and health-maintenance strategies.

**Table 1**
**LEVELS OF DETAIL IN FUNCTIONAL GENOMICS**

| Level of Analysis | Field of Study | Description | Status | Methods of Analysis |
|---|---|---|---|---|
| DNA | Genomics | Study of the complete set of genes of an organism or its organelles | Context independent | • Systematic DNA sequencing |
| RNA | Transcriptomics | Study of the complete set of messenger RNA molecules present in a cell, tissue, or organ | Context dependent, with the complement of messenger RNA varying with changes in physiology, development, or pathology | • Hybridization arrays<br>• Serial analysis of gene expression<br>• High-throughput Northern analysis |
| Protein | Proteomics | Study of the complete set of protein molecules present in a cell, tissue, or organ | Context dependent | • Two-dimensional gel electrophoresis<br>• Peptide mass fingerprinting<br>• Two-hybrid analysis |
| Metabolite | Metabolomics | Study of the complete set of metabolites (low-molecular-weight intermediates) in a cell, tissue, or organ | Context dependent | • Infrared spectroscopy<br>• Mass spectrometry<br>• Nuclear magnetic resonance spectrometry |

Source: *Nature;* SRI Consulting Business Intelligence (SRIC-BI)

## Model Genetic Systems

Because the study of nonhuman organisms will improve scientists' understanding of basic biological process that are relevant to human biology, a secondary goal of the HGP has been to map and sequence the genomes of model organisms. However, humans and model organisms differ not so much in the number of genes each carries but in the structure of genes and activities of their protein products. Exelixis (South San Francisco, California) has acquired a reputation as the premier model-system genetics-technology company. Although Exelixis formed to develop the platform on the fruit fly, the company has since acquired Agritope (Portland, Oregon) and Artemis (Cologne, Germany) to

bolster its access to other major genetic model organisms. Its main business services are to identify new drug targets and determine the mechanisms of action for existing drugs. Other model-organism companies include Zygogen (Atlanta, Georgia), Deltagen (Redwood City, California), Develogen (Gottingen, Germany), Union Biometrica (Somerville, Massachusetts), Devgen (Ghent, Belgium), and Ingenium Pharmaceuticals (Martinsreid, Germany).

In May 2002, the National Human Genome Research Institute (Bethesda, Maryland) selected six new "high-priority" organisms to sequence, basing the selection on scientific merit, particularly the potential human and economic benefit of studying each organism. The institute selected the chicken, chimpanzee, various strains of fungi, the honey bee, the sea urchin, and the eukaryotic protist *Tetrahymena thermophila*. It classified the protozoan *Oxytricha trifallix* and the rhesus macaque as "moderate-priority" organisms. The Institute for Genomics Research (Rockville, Maryland) has received a five-year contract for $25 million from the U.S. National Institute of Allergy and Infectious Diseases to establish a Pathogen Functional Genomics Resource Center. The center's initial goals are to select several target pathogens, particularly *Streptococcus pneumoniae*, *Staphylococcus aureus,* and *Salmonella typhimurium*, and begin preparing core research resources—such as specialized reagents, gene clones, and microarrays—for use by the research community.

Scientists have sought to sequence more than 150 microbial genomes since the completion of the first microbial genome sequence, *Haemophilus influenza*, in 1995. The U.S. Department of Energy's (DOE's) Microbial Genome Program plans to use sequenced microbial genomes to pursue opportunities in areas such as environmental-waste cleanup and energy production. The DOE's Genomes to Life program is attempting to develop a genome-based understanding of the components and processes of living systems, hoping to develop an integrated and predictive understanding of biological systems and to gain insight into how both microbial and human cells respond to environmental changes. In a complementary development, the Microbial Cell Project is studying the complexity of organisms at a systems level, particularly the function and regulation of genes for a single living system and the pathways in which the protein products interact.

## Gene-Expression Analysis

Gene-expression analysis is among the more mature emerging technology areas in genomics. DNA microarrays have been a critical enabling technology for expression analysis, particularly for the identification of candidate genes in drug-discovery research. The ability of DNA chips to look at a large number of genes simultaneously in a high-throughput manner has driven demand among researchers for such tools. Insights that have emerged from gene-expression analysis have been encouraging. Rosetta Inpharmatics (Kirkland, Washington) research indicates the potential to recognize aggressive tumours from the genes they express, which could ultimately help physicians diagnose various cancers more effectively and treat them more selectively than they can

now. Researchers are taking encouragement from their growing ability to decipher gene-expression profiles specific to diseases, and they see potential for the technique to improve disease management and outperform all currently used clinical parameters in predicting disease outcome.

The completed genome sequences of numerous organisms now allow researchers to combine technological development with emerging data sets. Global- or whole-genome analysis will be important in understanding the function of genes or the activities of compounds. The Affymetrix Human Genome U133 gene-chip set, which uses the publicly available draft of the human genome, comprises two microarrays of more than 1 million unique oligonucleotide features that represent the approximate number of human genes. This technology allows researchers, for the first time, to analyze the whole human genome on two chips. Because the set contains the largest number of data points ever on a commercial array, it allows researchers to assess gene-expression levels genomewide in a single experiment.

Although the development of gene-expression databases has also been a major focus in both the public and private sectors, much of the data generation in gene-expression analysis still takes place without commonly accepted standards and protocols. On the commercial front, Molecular Dynamics (now part of Amersham Biosciences) and Affymetrix formed the Genetic Analysis Technology Consortium in 1997 to standardize the rapidly growing field of array-based genetic analysis, providing a unified technology platform to design, process, read, and analyze DNA arrays. Rosetta Inpharmatics has also proposed standards for representing gene-expression data in the form of the Gene Expression Markup Language. But progress in both efforts has been slow, and compliance has been uneven. The Microarray Gene Expression Database group aims to use a protocol it calls MIAME (minimum information about a microarray experiment) to facilitate the standardization of DNA-array experiment annotation and data representation.

**Table 2**
**STANDARDS FOR GENE-EXPRESSION DATA**

| Name | Source | Standard |
|---|---|---|
| Genetic Analysis Technology Consortium | Affymetrix and Molecular Dynamics | Proprietary |
| Gene Expression Omnibus | National Center for Biotechnology Information | Tab Delimited Text |
| Minimum Information about a Microarray Experiment | European Bioinformatics Institute and Microarray Gene Expression Group Consortium | Extensible Markup Language |
| Gene Expression Markup Language | Rosetta Inpharmatics | Extensible Markup Language |
| GeneXML | GeneX Project at the National Center for Genome Resources | Extensible Markup Language |

Source: SRIC-BI

## Pharmacogenomics

Pharmacogenomics is the use of information about genetic variation that researchers derive through the study of the human genome to understand how individuals respond to therapeutics. Genetic variations, which can manifest themselves as differences in drug targets or as differences in the enzymes that metabolize drugs, can affect the therapeutic effectiveness or toxicity of a drug. In recent years, the focus of research has been gradually shifting from the discovery of single-nucleotide polymorphisms (SNPs) to identification of high-throughput, low-cost methods of assaying SNPs. SNPs are single-nucleotide variations in DNA sequences that can occur in both the coding and noncoding regions of the genome. More recently, haplotypes or sets of SNPs have come to prominence as a potentially more reliable means to correlate genetic variations with disease.

If researchers are to identify the SNPs that are most valuable as clinical markers for particular disease traits, the technologies for assaying SNPs will need to evolve, and scientists will need to agree on an optimal scoring strategy to identify statistically significant SNPs. Three experimental concepts vie for dominance in SNP genotyping today: allele discrimination (such as hybridization or enzymatic reactions), detection methods (such as fluorescence or mass spectrometry), and assay formats (such as chip-based or bead-based formats). Regardless of the underlying technology, a genotyping system for SNPs needs to be robust, cheap, multiplexed, and fully automated.

Haplotypes are sets of genetic markers that are close enough on a chromosome for a person to inherit them together. They elevate SNPs to an entirely new level of power because they enable scientists to identify common patterns of variation or blocks of genetic markers that occur throughout the population. This discovery will likely result in a faster, more simplified, and less costly approach to identifying genes responsible for disease. In the United States, the Center for Genome Research at the Whitehead Institute (Cambridge, Massachusetts) has been promoting a public map with 200 000 common genetic variations throughout the genome, stating that is sufficiently detailed for research purposes. In the private sector, several companies are pursuing proprietary haplotype strategies for identifying disease genes. These players include Gennaissance Pharmaceuticals (New Haven, Connecticut), Illumina (San Diego, California), Celera Genomics (Rockville, Maryland), and Affymetrix spin-off Perlegen Sciences (Mountain View, California).

## Proteomics

Proteomics embraces any large-scale characterization procedure for proteins. According to the Human Proteome Project (HUPO), proteomics work seeks to identify all the proteins encoded in the human genome and then to determine their range of expression across the cell types within human tissues, their subcellular localization, their posttranslational modifications, their interactions with other proteins, and their structure-function relationships. Efforts to date have focused on cataloging information. Over time, the work will expand to elucidate the dynamics of proteins and develop higher levels of biological knowledge by identifying and mapping protein interactions in metabolic or signaling pathways, for example.

The core separation, identification, and characterization technologies that find use in proteomics research are two-dimensional polyacrylamide gel electrophoresis and mass spectrometry. However, the absence of a high-throughput, automated method of identifying, characterizing, and quantifying proteins has been a major technical limitation in advancing the field. The field lacks an amplification technology for large-scale protein identification and analysis that is similar to the polymerase chain reaction for DNA. A new approach to quantitative proteomics uses a new class of chemical reagents, isotope-coded affinity tags (ICATs). Applied Biosystems (Foster City, California) has exclusive licensing rights to these ICAT reagents. Other technology platforms for protein identification and protein-interaction mapping include Myriad Proteomics's (Salt Lake City, Utah) ProSpec and Curagen's PathCalling, respectively. Database providers include Protein Pathways (Woodland Hills, California) and Axcell Biosciences (Newtown, Pennsylvania).

HUPO recently announced five initial research projects that reflect a mix of technology, tools, and research. Funding appears to be readily available to support these initial projects. For example, German researchers are requesting a £40 million grant from the European Union for the antibody initiative, and Korea's ten-year 21st Century Frontier Research Program is considering devoting some funds to Asian researchers'

effort to form a liver-proteome consortium. Meanwhile, the U.S. National Institutes of Health is still evaluating how it can best encourage the field's development. The five initial HUPO projects focus on:

- *Plasma proteome.* The project aims to identify less abundant proteins in blood, initially in healthy adults.
- *Antibodies.* The project team will build a library of antibodies for 30 000 gene products.
- *Cell models.* Scientists will implement a liver-proteome project and coordinate data standards for existing studies.
- *Bioinformatics.* The project team will develop databases, analysis software, and annotation standards.
- *New technology.* The project aims to develop methods of quantifying 5000 proteins and interactions in tissue/cell types.

## Chemogenomics

The combination of chemistry and genetics—chemical genetics or chemogenomics—is an emerging field in functional genomics. The principle is to use small molecules to screen targets so that scientists can identify key genes in a specific biological pathway of interest and investigate their role in disease. Chemical genomics also offers opportunities to understand poorly understood modes of action of well-known therapeutics and to identify potential drug leads and optimize promising drug candidates. The growing and widespread availability of technologies—from high-throughput screening and combinatorial chemistry to genomics information and array technologies—is facilitating activity in the field.

The U.S. National Cancer Institute (NCI) will establish an open-access database to store all available data on the effects of small chemical molecules on proteins, cell pathways, and tissue formation in the field of chemical genetics. Toward this end, the NCI will establish "molecular targeting laboratories" to synthesize and screen thousands of small molecules for their biological effects. Harvard University's Institute of Chemistry and Cell Biology has received a $40 million, five-year contract from the NCI to develop tools such as protein arrays and to build a public database that will eventually catalog as many as 1 million small molecules that block or interact with proteins.

Companies active in this area include ChemGeneX Therapeutics (Menlo Park, California), Infinity Pharmaceuticals (Boston, Massachusetts), NeoGenesis (Cambridge, Massachusetts), Morphochem (Munich, Germany), ComGenex (South San Francisco, California), Amphora Discovery (Research Triangle Park, North Carolina), and Ambit Biosciences (San Diego, California). Iconix Pharmaceuticals (Mountain View, California) is developing a chemical-genomics database, ChemExpress, to enable scientists to relate and connect disparate scientific data such as gene-expression profiles, chemical structures, and biological activities for novel drug-discovery insights. Iconix is using the database to develop a portfolio of proprietary drug signatures that provide the

molecular definition of a drug's or lead compound's in vivo activity and to predict potential toxicity and efficacy in pharmaceutical drug-optimization programs. Incyte Genomics (Palo Alto, California) has exclusive market and selling rights to ChemExpress.

## Toxicogenomics

Toxicogenomics—the fusion of functional genomics and molecular toxicology—is a growing field of activity in which researchers attempt to find correlations between responses to toxicants and changes in the genetic profiles of subjects who undergo exposure to such toxicants. Advances in gene-array technology are driving progress in toxicogenomics. Public agencies provide the majority of resources for microarray research of environment-gene interactions, and much of the research takes place at or through government laboratories. In the United States, participants include the National Institute of Environmental Health Sciences (NIEHS), the National Center for Toxicogenomics, the Chemical Industry Institute of Toxicology, the National Center for Toxicological Research, the Centers for Disease Control and Prevention's National Center for Environmental Health, the National Cancer Institute, and the Environmental Protection Agency Microarray Consortium.

In 2001, the NIEHS awarded five-year grants totaling more than $37 million to five U.S. academic research centers. The United States is the primary focus of much research activity. A growing number of U.S. initiatives focus on environmental health issues. An important commercial interest in toxicogenomics is to determine sequence polymorphism and gene-expression patterns among people with heightened sensitivity to the adverse effects of drugs. In theory, toxicogenomics differs from pharmacogenomics. The former identifies people who exhibit adverse responses to drugs, whereas the latter identifies the population subgroups that benefit the most from a particular drug. Commercial providers of microarrays, screening services, and related informational databases in the field of toxicogenomics include Xenometrix (San Diego, California), PHASE-1 Molecular Toxicology (Santa Fe, New Mexico), Affymetrix, Clontech (Palo Alto, California), and Gene Logic (Gaithersburg, Maryland).

## Epigenomics

The science of epigenomics studies how to alter DNA activity through chemical processes rather than genetic mutations. Epigenetic mechanisms that affect gene-regulating activity include DNA methylation and RNA interference; these mechanisms can persist through one or more generations. For example, the addition of a methyl group to cytosine, one of the four nucleotide bases that make up DNA, appears to play an important role in cell regulation of gene expression. Companies are trying to put into practice the insights they've gleaned from epigenetics, particularly in disease diagnostics and drug development. The Human Epigenome Consortium formed in 1999 through the participation of the Sanger Centre, Centre National de Génotypage in France, the German Cancer Research Centre, the Technical University of Berlin, the Max-Planck-Institut for Molecular Genetics, and Epigenomics (Berlin, Germany).

Epigenomics is studying methylation patterns in different tissue types to identify important mechanisms of disease that might otherwise not be evident through traditional genomics approaches. Epigenomics's ability to use such information to generate so-called digital phenotypes will enable it to build a personalized-medicine business around a DNA-methylation platform. Sangamo BioSciences (Richmond, California) is targeting cancer in its epigenetic-focused drug-discovery and -development activities. With the ability to alter epigenetic status to upregulate the expression of a needed gene or a gene that will counter the effect of an unneeded gene, researchers are confident that epigenetics has a worthwhile role to play in the future.

## Metabolomics

Scientists describe mutations that do not produce an obvious phenotype as "silent mutations." These mutations suggest that an organism has mechanisms that allow it to compensate for the loss of a gene, for example by rerouting through alternative pathways for the same desired end product or effect. Metabolomics, which seeks to delineate all metabolic processes in an organism, aims to correlate metabolic reactions with the presence or absence of individual genes. Metabolomics could also ultimately lead to the identification of optimal genetic combinations for particular metabolic routes, a finding that would have implications for plant breeding and transgenic organisms that make specific compounds, for example.

Unlike proteins and genes, the range of biochemicals that are available for analysis is not chemically homogeneous and spans a disparate range of compounds. Moreover, many of the technological approaches in metabolomics are at an early stage of development. Researchers at the Universities of Manchester, Cambridge, Wales, and Amsterdam and at Amsterdam Free University are collaborating in a research project to use yeast to develop a novel method—functional analysis by co-responses in yeast—for measuring all low-molecular-weight components. Ronald Breaker and colleagues at Yale University have developed a chip-based technique for distinguishing and identifying metabolites that uses RNA switches they have deposited on a gold-coated silicon plate.

## Systems Biology

The key challenge in molecular biology is to study how component parts function collectively as complex biological systems. Systems biology calls for defining the elements of a system, determining what relationships exist between those elements, and developing models that accurately represent the structure and behavior of the system. Achieving such an integrative or systems-biology approach will require continuing convergence of molecular biology, computer science, and the physical sciences. This systems approach requires a reiterative research process in which researchers start with a theoretical model of a biological system, conduct experimental research using that theoretical understanding, and use the resulting data to build more accurate models. The goal is to develop a model that will be more predictive of the structure and behavior of certain component parts than current methods are.

Integration of information technology with biotechnology is critical for the field to advance, particularly to support data mining, warehousing, and integration as well as the development of models. David Haussler, a leading bioinformatician, states that in the future, research specialists will need to be able to tie into a common infrastructure of knowledge, data, visualization, analysis by algorithm, model by machine, and simulation *in silico* and that the task of bioinformatics is to build this infrastructure. Leroy Hood, a leading proponent of systems biology, founded the Institute for Systems Biology (Seattle, Washington) to pioneer interdisciplinary research in the field and promote cross-fertilization in science. Academic institutions have also been at the fore of systems biology, with initiatives at the California Institute of Technology (Biological Sciences Initiative), the University of California, Berkeley (Health Sciences Initiative), and Stanford University (Bio-X). The University of Chicago, Harvard University, and Princeton University also plan to establish new interdisciplinary ventures. Among the companies active in systems biology are Molecular Mining (Ontario, Canada), MelTec (Magdeburg, Germany), Beyond Genomics (Waltham, Massachusetts), and Zygogen.

## BIOINFORMATICS

As the science and technology evolve in life-sciences research, so do the concept, definition, and business model of—and marketplace for—bioinformatics. The application of informatics in life-sciences research creates problems that stem partly from the scientific process and partly from the enabling technologies. For end users, the bioinformatics challenge is how to overcome the problems of too much data, isolated research, and fragmented communication. Moreover, biological data are highly heterogeneous in format, organization, quality, and content. In turn, the fast pace of change in the research field creates uncertainty about the ability to provide lasting solutions, and the lack of standardization in research tools and techniques creates technology islands that often generate incompatible data sets.

Though none of the technological developments or advances in genomics would be possible without the underlying information technology, the actual process of life-sciences research also acts as a catalyst for new directions and models in high-performance computing. For example, among the experimental computing solutions that have provided greater processing capability and distributed tasks in biology are dedicated and managed networked processors (Linux clusters) and distributed computing resources (grid computing). Bioinformatics is increasingly a source of competitive advantage as data integration and knowledge discovery become the bottlenecks in the current drug-discovery process. Solutions are available from industry stalwarts LION Bioscience (Heidelberg, Germany) and Compugen (Tel Aviv, Israel) and are starting to emerge from start-ups such as Anvil Informatics (Burlington, Massachusetts), Incellico (Durham, North Carolina), GeneticXchange (Menlo Park, California), and Confirmant (Oxfordshire, England). Biospice, which owes its start to the U.S. Defense Advanced Research Projects Agency, is an open-source framework and toolkit for developing computational biology models for representing and simulating cellular processes. (Further information is available online at http://www.biospice.org/ .)

Open standards from the Interoperable Informatics Infrastructure Consortium (I3C), which define a common interoperability framework for the exchange and management of data from a variety of sources and across the entire life-sciences community, are starting to emerge. The I3C wants to leverage the emerging technology of Web services, which airlines, Web sites, and phone companies already use to link parts of different companies' data infrastructures. Also under development is an integration platform, Omnigene (http://omnigene.sourceforge.net ), and Integr8, an integration project of the European Bioinformatics Institute. The problem of information integration will likely remain until genomics data become more stable. The ideal solution is first to develop standardized or reference data sets of the human genome sequence.

## OUTLOOK

The field of genomics will gain momentum as a result of three key factors: researchers' growing recognition that biology is an information science, the emergence of high-throughput biology, and awareness of the fundamental role that applied mathematics and computer science play. By stimulating such changes, the HGP is the catalyst transforming biology and medicine. As the emphasis in research shifts toward the field of functional genomics, new theories will emerge about biological systems and will change our view of the basic processes of life as well as provide new understanding. This new understanding will in turn give rise to improved methods of disease diagnosis, prevention, and treatment.

Many unknowns accompany this process, however. They include uncertainty about the rate and direction of actual knowledge discovery in the science of genomics, issues about ownership rights to genetic resources, and the ethical, social, and legal implications associated with modern science and its application. As a result, the science has to become more sophisticated to advance knowledge, the industry needs better rules of operation to function effectively, and future commercial viability and market adoption depend on improvements in scientific literacy among the general public.

**FOR ADDITIONAL INFORMATION** about the topics in this study, submit inquiries through your Executive Contact to the Business Intelligence Program (B-I-P) in your area (see the addresses opposite).

**SRIC-BI's BUSINESS INTELLIGENCE PROGRAM** provides clients with a system of integrated research intelligence and customized services to help define new market opportunities, identify and communicate future challenges, formulate and implement business strategy, and develop innovative products, processes, and services. Clients access the following research and consulting services:

- *Scan: Signs of Change*. *Scan* provides an early warning of emerging trends that could present threats or opportunities to B-I-P clients.

- *Research Reports.* Throughout the year, B-I-P issues detailed reports about defining forces driving change in today's business environment. These reports provide clients with an assessment of the impacts of the defining forces—and what they mean for business.

- *Insights.* Announced in the *Bulletin* and available on request, these publications provide additional coverage of defining forces and other developments in technology, market, and management topics.

- *Bulletin.* B-I-P's monthly *Bulletin* highlights *Scan*, new Defining Forces reports, and the on-request *Insights* available to clients. It announces special B-I-P and SRI activities.

- *Consultation Services.* Clients may submit inquiries and consult with SRIC-BI professional staff to address specific market questions, technology issues, or management process concerns. For telephone inquiries to Menlo Park, call the B-I-P Inquiry Line: +1 650 859 4600; fax: +1 650 859 4544; or e-mail inquiry@sric-bi.com.

- *Forums and Workshops.* Clients may join SRIC-BI researchers, industry pioneers, and other clients for roundtable discussions and business workshops to explore new business trends and technology topics in more detail.

- *Web Access.* B-I-P materials are available via the Internet.

For further information about the Business Intelligence Program, consult your Executive Contact or write to:

SRI CONSULTING BUSINESS INTELLIGENCE
333 Ravenswood Avenue
Menlo Park, California 94025
Telephone: +1 650 859 4600
Fax: +1 650 859 4544

SRI CONSULTING BUSINESS INTELLIGENCE
Knollys House
17 Addiscombe Road
Croydon, Surrey
CR0 6SR, England
Telephone: +44 20 8686 5555
Fax: +44 20 8760 0635

SRI CONSULTING BUSINESS INTELLIGENCE
Daito Building 2F.
7-1, Kasumigaseki 3-chome
Chiyoda-ku, Tokyo 100, Japan
Telephone: +81 3 5251 1747
Fax: +81 3 5251 1754

Visit our Web site at http://www.sric-bi.com .