Protecting bioinformatics' value

As the field matures, so too must intellectual property strategies to capitalize on bioinformatics' wealth.

BY MICHAEL A. GOLLIN

Bioinformatics—the use of computers to characterize the molecular components of living things and other biological data—is still a youngster in the life sciences family. By combining the revolutionary tools of information technology with biotechnology, bioinformatics channels the flood of information pouring from genomics and proteomics laboratories and promises to accelerate the discovery of new diagnostics and therapeutics.

But bioinformatics is growing up, and intellectual property management strategies need to grow with it. The first generation of bioinformatics ended in 2001 with the sequencing of the human genome by Celera Genomics and the Human Genome Project international consortium, which proved the power of the new tools. The U.S. Patent and Trademark Office organized an examining group to deal exclusively with bioinformatics, signaling its acceptance of this discrete technology field.

Now, bioinformatics companies are either maturing from purveyors of data into vertically integrated drug discovery enterprises, or expanding their role as service providers. Several examples prove the point.

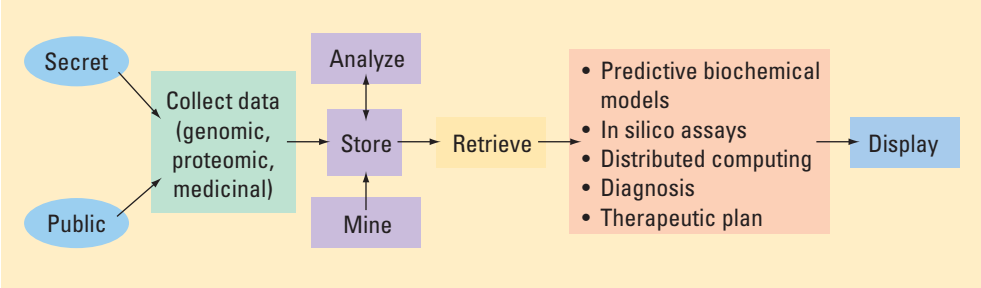
In 2002, Celera shifted away from its emphasis on gene sequence database services and focused instead on discovering therapeutics and diagnostics. The company was integrated more closely with its affiliate, Applied Biosystems, which sells research instruments, reagents, and software, and their joint venture, Celera Diagnostics. Celera's database of annotated genetic sequence information is now one part of a larger effort to find concrete products for research and medicine.

Likewise, citing competition from public domain genomic information sources, Incyte has closed its California bioinformatics research facility and shifted to a drug discovery approach. The company has about 100 patents relating to bioinformatics (microarray fabrication, measurement of gene expression, RNA amplification, sequencing, model organism mutagenesis, and library construction). But that number is a small fraction of its nearly 800 patents, most of which cover specific disease-related genes. For both Celera and Incyte, bioinformatics has been an invention engine that churns out new patentable genes and related biomolecules, rather than a business end in itself.

Meanwhile, Gene Logic retains its focus on bioinformatics services to other companies, but in 2003, the company acquired clinical research company TherImmune, bolstering its gene expression data production, management, and software with TherImmune's drug development services (toxicology, pharmacology, chemistry, and clinical trial management). In doing so, Gene Logic has joined a new generation of bioinformatics firms integrating genomic and proteomic data with the downstream drug development process, including clinical research.

Finally, bioinformatics systems are being applied to integrate research data, medical records, and other life sciences information, giving rise to the new terms "clinomics" and "theragnostics." This emphasis on electronic databases has, for example, led IBM to refer to its life sciences initiatives as "information-based medicine," and has even risen to the level of presidential politics, with both candidates calling for improved health care data management systems.

As we enter this integrated second generation of the bioinformatics revolution, where is the greatest value of the technology, and how can bioinformatics researchers identify and capture that value? From an intellectual property perspective, the best approach is to tease apart the various innovative and valuable components of a particular technology and to use whichever tools—patents, trade secrets, trademarks, and copyrights—are available to offset inherent weaknesses in each type of intellectual property.



Bioinformatics involves many data interactions and components, all of which offer opportunities for innovation and possible intellectual property protection.

Intellectual property tools

It is a fundamental tenet of modern industrial society that intellectual property laws advance technology by providing incentives and rewards to creative people, who can take exclusive rights over the fruits of their labors. The tools of intellectual property protection can capture the value bioinformatics brings to the life sciences. Once captured, these tools can be used to spur investment and further research and development, which are necessary if bioinformatics is to achieve the lofty goals of its second generation.

People working in bioinformatics use and produce a wide range of innovations: computer hardware and software; associated analytical biological instrumentation, systems, and

components; biological and chemical reagents; and massive databases and methods of using them, including Internet-based subscriptions and other new business models. Figure 1 summarizes the interactions and components of data in bioinformatics. Innovation can be found in each of the components, and in the interactions between them.

Using the appropriate intellectual property tools in an integrated way, innovators can assert a wide variety of exclusive rights, helping recoup their research investments. These tools include the following items.

Trade secrets. Those who obtain secret information may protect it as trade secrets by entering into confidentiality agreements to limit access and use. Trade secrets may be found in proprietary database architecture, the data itself, and the database analysis scripts that query the data. The scripts themselves may have the most value, and they can be maintained as trade secrets, with access only by subscription or other contract. Such contracts can prohibit reverse engineering. One weakness of trade secrecy is that it does not prevent independent production of the database and tools. Also, publication, whether accidental or intentional, eliminates trade secrecy.

Copyrights. Software can be protected from outright copying, as can data arrangements such as annotated genome sequences. However, copyright does not prevent anyone from writing a new program that performs the same analysis. Likewise, copyright does not protect raw data.

U.S. copyright law, per se, does not protect databases. European Directive 96/9/EC on the legal protection of databases protects the maker who has made a "substantial investment in obtaining, verifying, or presenting the contents." For 15 years from the creation or significant update of the database, no one may extract or reuse all or a substantial part of the database contents.

Trademarks. Bioinformatics companies can use trademarks to capture the value of the known quality of their services and data. As noted, Celera and Gene Logic became recognized leaders, and researchers are likely to favor their new products over those of lesser-known companies. Companies with established brands may also find it easier to obtain financing and strategic alliances. But this is a very sophisticated marketplace, unlike consumer products, and trademarks provide only minimal competitive advantage.

Patents. A patent gives unrivaled control over the right to use an invention, with powerful judicial remedies that complement weaknesses in other types of intellectual property protection. However, patents create a tension between the desire to keep information secret and the requirement for full disclosure in a patent application. Bioinformatics is a particularly tricky area, requiring familiarity both with the precedents and guidelines for computer and software inventions, and the equally complex rules for biotechnology inventions. Moreover, international rules about software and biotechnology are even stricter than in the United States. Accordingly, bioinformatics patents require extra effort, and thus cost, to obtain, and are at higher risk for challenge if enforced.

The organizations that will succeed in moving bioinformatics into the mainstream will have teams of scientists, business people, and attorneys who can work together.

Claiming the value with patents

All of the computational and statistical tools shown in Figure 1 are patentable if they satisfy the criteria of patentability (novelty, nonobviousness, and so on). These include tools used to analyze genomic and proteomic information, as in collecting, storing, analyzing, and retrieving data; data mining and data visualization tools; sequence alignment and pattern recognition tools; molecular modeling tools; and predictive tools.

It is one thing to know what types of inventions are patentable, and yet another to know which patent claims are the most valuable and enforceable. A few general pointers may be helpful.

First, it is crucial that the claimed method, software, system, or the like will be practiced by an individual, against whom an enforcement proceeding may be pursued. As practitioners of Internet patents know, it is best not to claim a distributed computing "system" as a whole, because it would be infringed only by a group of people. Claims to the server side of the system are best in this situation.

Second, given uncertainties in bioinformatics technology and the relevant law, claim diversity is important. Thus, information technology claims can be drawn to a hardware system, software running on it, subsystems, components, and data compiled or assembled using the system. Furthermore, claims can be drawn to methods of producing the components, data, and associated analytical instrumentation, as well as methods of operating and using them. In addition, biotechnology claims can be drawn to the genes, proteins, sequences, fragments, vectors, and host cells identified and/or isolated using bioinformatics tools. Methods of obtaining and using these biological materials are also patentable. Thus, one innovation may result in multiple patents with many claims drawn to different aspects, an increasingly expensive approach. Hence, more effort is put into patenting the fruits of bioinformatics-driven research (the genes and proteins and small molecules that interact with them) than into the bioinformatics systems themselves.

Third, claims should focus on the most valuable aspects of the innovation. A patent on a gene chip would capture the value of the chip, but a claim covering the overall system with analytical instrumentation and computational readout components for the inventive chip would have much higher value and establish a larger royalty base. Also, the database formed by the biochip and system may ultimately have the highest value. Although data is not patentable, the architecture or "data structure" can be. The data structure can be claimed in "product-by-process" format, for example.

Finally, some of the early revenue-generation models—for example, e-commerce-based data subscriptions—were inventive and susceptible to protection (in the United States) as business methods, like the Amazon one-click patent and the State Street Bank financial services model.

It is beyond the scope of this article to analyze specific bioinformatics patent claims.

Many examples are available for review on public databases, as are examples provided in training materials from the U.S. Patent & Trademark Office (see "Further reading"). Most importantly, inventors and business people need to consult with a patent attorney for advice with respect to their particular technology.

Working together

Generally speaking, success in innovation requires good management of technical, business, and legal skills. Thus, scientists, business people, and lawyers need to work together in teams. As bioinformatics becomes pervasive throughout the pharmaceutical innovation life cycle and into the therapeutic stage, these teams will include more than just the hybrid biotechnology–computer specialists who created the field.

The broader groups to be involved include molecular and cellular biologists, physicians, clinical researchers, and regulatory affairs specialists. The computer software and hardware industries have also poured resources into this market, and so electrical engineers and software writers should understand the challenges and opportunities. Anyone involved in providing service to these sectors should also know the basics. This includes reagent providers, research service organizations, and professionals, such as patent attorneys and licensing specialists, as well as business consultants and accountants.

Doctors, hospitals, insurers, and patients have a stake in the issues as well. Investors, patent examiners, and government regulators round out the list of people who need to understand bioinformatics. These people work in public, academic, and corporate research organizations, including small biotech companies, large pharmaceutical companies, and generic companies; in government agencies; in hospitals; and in private health care industry companies.

The organizations that will succeed in moving bioinformatics into the mainstream will have teams of scientists, business people, and attorneys who can work together, first to identify the most valuable aspects of these new tools, and second to capture and protect those aspects with all the intellectual property techniques available. Those who invest the required time and money will be rewarded with the ability to bring whole new types of information, drugs, crops, and services to the general public.

Further reading

Eyler, Y. Proteomics Examination; www.uspto.gov/web/patents/biochempharm/documents/bcpmar02.pps.

Gollin, M. Managing the cost of intellectual property. *California Lawyer,* Nov 2002.

Woodward, M. P. Bioinformatics and §101; www.uspto.gov/web/patents/biochempharm/documents/woodward.pps.

Woodward, M. P. SniPpetS and Bytes; www.uspto.gov/web/patents/biochempharm/documents/snippets.pps.

Written description guidelines, example 12; www.uspto.gov/web/menu/written.pdf.

Michael A. Gollin is a patent attorney with Venable LLP, based in Washington, DC. ■