BROWN RAYSMAN

BROWN RAYSMAN MILLSTEIN FELDER & STEINER LLP

BIOINFORMATICS 101: BASICS FOR THE COMPUTER LAWYER

by Richard Raysman
and
Peter Brown*

At the intersection of information technology and biology lies Bioinformatics. Simply put, Bioinformatics is the use of computers to store, retrieve, analyze or predict the sequence or structure of biological information. It combines aspects of mathematics, statistics, computer science, and biology within a given set of partly discovered biological rules and IT solutions.

The incredible amounts of information derived from mapping genes (not only human genes, but also those of plants, animals, and microorganisms) coupled with the search for new cures and ways to prevent diseases has revealed an insufficient information processing infrastructure at drug companies. Next generation networks, customized software, innovative database tools, new database storage methods and improved computational techniques are all required to determine and analyze recently discovered genetic information. As a result, labs and drug companies have been beefing up their IT staff and looking for IT support from computer and software companies. Because IT companies are scrambling to meet the needs of drug companies, IT attorneys should stand up and take notice. But before diving into Bioinformatics agreements, the IT attorney should first have a fundamental understanding of the scientific elements involved.

Biology. In the nucleus of any given cell are chromosomes, which contain DNA strands of genetic information (life's instruction manual). The most basic unit in a DNA strand are called nucleotides. Each nucleotide consists of an organic base and other elements not pertinent to this discussion. All species have the same four types of bases commonly known as A, G, C, and T. Bases in nucleotides paired together in various sequences help to form the double helix structure of a DNA strand. Human beings have about 3 billion pairs of bases. A single human gene may vary from a string of a few hundred base pairs to a million base pairs. The process of determining the sequence of bases in a single gene is called gene sequencing.

The human genome refers to all the genes in the human body, which is estimated to have about 30,000 genes. Mapping the human genome (or any other species' genome) requires identifying and locating specific genes, analyzing the position of the gene with other genes, and calculating the position of the gene relative to the ends and center of the chromosome. This process is called gene mapping.

From the genes' codes, proteins are produced, which in turn ultimately have a significant effect on a wide variety of characteristics of living organisms. Specifically, protein function is derived from specific instructions from a DNA base sequence. Like a piece of a jigsaw puzzle, every proteins' function is critical in the big picture of how they

interact with other proteins and chromosomes. If a protein functions incorrectly because of a mutation in the DNA base sequence, a disease such as cystic fibrosis or breast cancer could be the result. There are one million or so proteins in the human body. When attempting to identify all the proteins by analyzing their shapes and how they interact in the body coupled with the information derived from gene sequencing and gene mapping, it becomes apparent that studying the genome generates phenomenal amounts of information.

IT Solutions. Whenever large amounts of information need to be stored, analyzed, and manipulated, use of computers and information technology is invariably in the picture. In computers worldwide, are vast amounts of digital biological information. Some analysts have characterized the amount of genetic information derived from studying the human (and other species) genome not in gigabytes or terrabytes, but in petabytes (quadrillions of bytes).

To put the practical IT aspects of Bioinformatics into perspective, Celera Genomics Corp. is a good example. To map the human genome, two-thirds of Celera's staff are software engineers rather than scientists. Celera has also partnered with Compaq Computer Corp. to build a cluster of computers capable of processing 100 trillion operations per second, 500,000 times faster than a personal computer.

To store and access the genetic information as a future resource, labs and drug companies rely heavily on a customized IT infrastructure. Specifically, researchers and scientist require new computational techniques, customized software, and efficient database management systems. However, it is one thing to sequence, map, identify and store genetic information and quite another to analyze, process and determine how the genes affect the shapes of proteins and how they work in the body. The science of Bioinformatics includes this second step of analyzing the recently discovered information derived from gene mapping and gene sequencing. Indeed it is this next level that can lead to new discoveries and possibilities in gene therapy (the process of preventing and treating diseases).

Analyzing stored genetic data requires interaction with biological databases, which entails browsing, finding information about specific portions of a gene, executing computational queries, and organizing vast networks of information. To conduct such complex analyses and to cull through the plethora of nucleotide sequences, protein data and other genetic information, researchers and scientist rely on sophisticated data mining tools. Use of intricate search algorithms and statistical analysis are required to match the stored genetic information to any given query.

By some accounts, IT solutions have been credited for shortening the typical fourteen year drug-development cycle by four or five years and have cut costs by at least ten percent. With the recent mapping of the human genome and the requirement of sophisticated IT solutions, tech giants such as IBM, Sun and Oracle have been scrambling to fill the needs of drug companies. Market analysts have predicted that when high-performance computing combine with sophisticated data management systems, the global market for IT products and services in the biotech sector will grow to $43 billion by 2004.

This rapidly developing field requires IT attorneys to be prepared for the legal issues that may arise when involved in any Bioinformatics transaction.

Legal Implications in Bioinformatics

When negotiating a software transaction, IT attorneys typically focus on such provisions as acceptance testing, service levels, upgrades and enhancements, confidentiality, ownership and licensing rights between the parties. However, because of the nature of Bioinformatics, IT attorneys should also be aware of (i) a potential third party's intellectual property rights in patentable algorithms and (ii) the government's rights in government-funded inventions.

Patenting Algorithms. Oddly, the patent case that may have a substantial impact in the biotech industry does not involve biological science at all. Rather the focus is on mathematical algorithms. Since State Street Bank & Trust Co. v. Signature Financial Group Inc. and its progeny, the Federal Circuit Court of Appeals ruled that mathematical algorithms are patentable if they produce "a useful, concrete and tangible result."[1] The court went on to state that, "the mere fact that a claimed invention involves inputting numbers, calculating numbers, outputting numbers, and storing numbers, in and of itself, would not render it nonstatutory subject matter unless, of course, its operation does not produce a useful, concrete and tangible result."[2]

Therefore, under this line of reasoning, algorithms, processes, or computational techniques employed to mine genetic information from a database are arguably patentable. Similarly, patent rights may arise from algorithms used in a software application to calculate protein and genetic mutations, sequences, shapes, locations and functions. Because Bioinformatics relies so heavily on computational methods, drug companies, labs, and software vendors may attempt to patent these types of inventions to create new revenue streams. The tools for Bioinformatics research may prove to be as valuable as the drugs which result from the research.

Such new patents, based on mathematical algorithms, require IT attorneys to be conscience of the possibility that their client may be infringing on a third party's intellectual property rights. As a result, in the case of representing a lab or drug company, IT attorneys should perform some due diligence by investigating a software vendor's use and ownership rights in the software applications it intends to license. The IT attorney should also keep in mind other contractual protections such as representations, warranties, indemnifications, and limitation on liabilities when it pertains to any software at issue.

Governmental Strings Attached. Prior to 1980, title to patentable inventions that were funded (wholly or partly) by the federal government remained with the government. With the enactment of the Bayh-Dole University and Small Business Patent Act of 1980[3] and several similar enactments regarding technology transfers[4], nonprofit organizations, small businesses, and those who contract with a federally funded lab are now permitted to retain ownership or exclusive license rights of most government-funded inventions. However, such rights are not without some restrictions.

In exchange for giving up its exclusive ownership rights, the government receives (i) (a) under the Bayh-Dole Act, a nonexclusive, nontransferable, irrevocable, paid-up,

---

[1] 149 F.3d 1368, 1373 (Fed. Cir. 1998), cert. denied, 525 U.S. 1093 (1999).
[2] Id. at 1374.
[3] 35 U.S.C. §§ 200-212.
[4] Stevenson-Wydler Technology Innovation Act of 1980 (15 U.S.C. §§ 3701-3717); Federal Technology Transfer Act of 1986 (15 U.S.C. §§ 3701-3714); National Competitiveness Technology Transfer Act of 1989 (15 U.S.C. §§ 3701-3710); and National Technology Transfer and Advancement Act of 1995 (codified in scattered sections of 15 U.S.C.).

worldwide license to practice the invention on behalf of the United States[5] or (b) under the technology transfer acts, the government retains ownership and grants an exclusive license for reasonable compensation;[6] (ii) a right to require that products made through the use of the invention to be substantially manufactured in the United States;[7] (iii) a right to prohibit future assignments of patent rights without first approving the assignment;[8] (iv) a portion of any royalty fees;[9] (v) a right to grant licenses to small businesses;[10] and (vi) march-in rights[11].

March-in rights are a form of compulsory license whereby the U.S. government, under certain circumstances, can force entities that received government funds to grant licenses on reasonable terms to third parties. Alternatively, the government can grant the licenses itself. The circumstances which gives rise to the government's march-in rights are (i) when the inventor(s) have failed to take effective steps in a reasonable amount of time to achieve a practical application of the invention, (ii) when the action is necessary for public health and safety reasons, or (iii) when the action is required by public use regulations.[12]

Such governmental rights mean that IT attorneys should consider whether a party to the transaction or a third party is wholly or partly funded by Federal grants or whether such parties contracted with a federally funded lab.

Conclusion

Because current biological research relies so heavily on computing, IT attorneys will find themselves representing clients in Bioinformatics agreements. Rather than diving into the transaction as if it were merely another software license, they must consider the unique history and needs of Bioinformatics transactions.

* Leslie Gladstone Restaino, Of Counsel to the firm, and Stanley U. Paylago, an associate at the firm, assisted with the preparation of this article.

This article appeared in the Tuesday, May 8, 2001, edition of the New York Law Journal.

[5] 35 U.S.C. § 202(c)(4).
[6] 15 U.S.C. § 3710a (b)(1).
[7] 35 U.S.C. § 204.
[8] 35 U.S.C. § 202(c)(7)(A).
[9] 35 U.S.C. § 202(c)(7)(B). Although the Bayh-Dole Act only grants royalty rights, the royalty scheme is provided for in the technology transfer acts at 15 U.S.C. § 3710c(a)(1)(A)(i).
[10] 35 U.S.C. § 202(c)(7)(D).
[11] 35 U.S.C. § 203.
[12] Id.