Intellectual Property at the Public-Private Divide: The Case of Large-Scale cDNA Sequencing

Rebecca S. Eisenberg
University of Michigan Law School, rse@umich.edu

Follow this and additional works at: http://repository.law.umich.edu/articles
Part of the Business Organizations Law Commons, Intellectual Property Law Commons, and the Science and Technology Law Commons

Recommended Citation
Intellectual Property at the Public-Private Divide: The Case of Large-Scale cDNA Sequencing

REBECCA S. EISENBERG

The Human Genome Project provides fertile ground for studying the role of intellectual property at the wavering boundary between public and private research science. It involves a major commitment of both public and private research funds in an area that is of significant interest both to research scientists working in university and government laboratories and to commercial firms. It thus provides a wealth of new scientific discoveries that are simultaneously potential candidates for commercial development and inputs into further research. Its obvious implications for human health raise the stakes of getting the balance between private property and public access right, particularly at a time when public attention is riveted upon the rising costs of health care. It profoundly affects the interests of the young biotechnology industry as well as the more established pharmaceutical industry, and thus offers an opportunity to compare the perspectives of two very different types of commercial firms. Moreover, intellectual property issues have been unusually conspicuous in the recent history of advances in the emerging field of "genomics," even by the standards of the patent-weary genetics and molecular biology communities.

© 1996 Rebecca S. Eisenberg. All rights reserved.

Rebecca S. Eisenberg is a Professor of Law at the University of Michigan Law School. She gratefully acknowledges helpful comments and suggestions from Bob Cook-Deegan, Heidi Feldman, Roberta Morris, and workshop participants at the University of Chicago Law School Roundtable's symposium on the Human Genome Project, the University of Michigan, and the Green Center for the Study of Science and Society at the University of Texas at Dallas. This research has been supported by the Office of Health and Environmental Research of the U.S. Department of Energy, the Cook Fund of the University of Michigan Law School, and the Green Center for the Study of Science and Society at the University of Texas at Dallas.
I. EST Patent Controversy

Controversy has been particularly acute over intellectual property rights in the results of large-scale complementary DNA (cDNA) sequencing, a technique that focuses on the small fraction of the genome that cells "express" in the form of proteins.¹ The Human Genome Project has as its ultimate goal the mapping and sequencing of the entire human genome.² Yet only an estimated three to five percent of the human genome contains the code for proteins that cells make and use in performing their biological functions.³ Through the use of enzymes, it is possible to make copies of the coding regions within the genome that a cell is in the process of expressing, in effect creating a heavily redacted version of the genome that eliminates all of the "junk DNA."⁴ These redacted DNA sequences are called complementary DNA ("cDNA") sequences. Because cDNA sequences code for proteins that are used by human cells, this approach allows for the relatively expeditious identification of DNA sequences that are particularly likely to have medical, and perhaps commercial, significance.

It was somewhat controversial in the early years of the Human Genome Project whether it made sense to allocate funds to large-scale cDNA sequencing with existing resources and technology.⁵ One scientist who thought this approach promising was Dr. Craig Venter of the National Institute of Neurological Diseases and Stroke.⁶ Working with randomly selected cDNAs derived from human brain tissue, Dr. Venter and his colleagues used automated DNA sequencing machines to obtain a DNA sequence for a small portion of each cDNA.⁷ Each of these partial sequences, called an "expressed sequence tag" ("EST"), is long enough to supply a unique identification for the gene from which it derives, but short enough to permit rapid partial sequencing of a large number of genes. Any given EST can then be used as a probe to find its corresponding full-length cDNA for further study. A database of ESTs thus provides a catalogue of expressed genes that can serve as a useful resource in subsequent research to identify particular genes of interest and to study their biological functions.

ESTs quickly became the focal point for a raging controversy over intellec-

³. Adams, 252 Science at 1651 (cited in note 1).
⁶. See Leslie Roberts, Gambling on a Shortcut to Genome Sequencing, 252 Sci 1618 (1991); Genome Investigator Craig Venter Reflects On Turbulent Past And Future Ambitions, Scientist 1, 10 (July 24, 1995).
⁷. Adams, 252 Science at 1651 (cited in note 1).
tual property rights in the human genome. The National Institutes of Health ("NIH") set off the debate when it began filing patent applications on the first few thousand ESTs from Dr. Venter's laboratory in 1991. The next year Dr. Venter and his group left NIH to form a nonprofit research organization, The Institute for Genomic Research ("TIGR"), with generous private sector funding. The same financial backers also formed a for-profit company, Human Genome Sciences ("HGS"), to identify and develop commercial products out of the sequence information developed by the nonprofit TIGR. The two organizations were soon both engaged in a massive automated cDNA sequencing effort, creating two large, privately-held databases of cDNA sequence information. Meanwhile, in California, another private firm, Incyte Pharmaceuticals, had also turned its attention to large-scale sequencing of cDNA fragments, creating a competing private database. Dismayed by the limited availability of these privately controlled databases to researchers, within two years a major pharmaceutical firm, Merck & Co., announced its entry into the field as sponsor of a competing cDNA sequencing effort at Washington University, for instantaneous dedication to the public domain.

Each of these developments has been met with lively speculation about its strategic significance from an intellectual property perspective. Are cDNA fragments of unknown function patentable, or is further research or characterization necessary before they satisfy patent law standards? Will patents on


such fragments promote commercial investment in product development,\textsuperscript{15} or will they interfere with scientific communication and collaboration and retard the overall research effort?\textsuperscript{16} In the absence of patent rights, how may the owners of private cDNA sequence databases earn a return on their investment while still permitting other investigators to obtain access to the information on reasonable terms?\textsuperscript{17} What are the rights of those who contribute resources such as cDNA libraries that are used to create the databases,\textsuperscript{18} and of those who identify sequences of interest out of the morass of information in the databases by formulating appropriate queries?\textsuperscript{19} Will the disclosure of ESTs in the public domain preclude patenting of subsequently characterized full-length genes and gene products?\textsuperscript{20} And why would a commercial firm invest its own resources in generating an EST database for the public domain?\textsuperscript{21}

Two factors have contributed to the fascination with intellectual property issues in this setting. First, there is a perception that some pioneers in large-scale cDNA sequencing have sought to claim intellectual property rights that reach far beyond their actual achievements to cover future discoveries yet to be made by other researchers.\textsuperscript{22} For example, the controversial NIH patent applications claimed not only the ESTs that were actually set forth in the specifications, but also the full-length cDNAs that might be obtained by using

\begin{thebibliography}{9}

\bibitem{Adler} See Adler, 257 Science at 910-11 (cited in note 14); Paley, 44 Syracuse L Rev at 1008-09 (cited in note 14).

\bibitem{Docherty} Docherty, 26 Akron L Rev at 547-48 (cited in note 14); McKay, 10 Santa Clara Computer & High Tech L J at 490 (cited in note 14); Smith and Kettelberger, 22 Amer Intell Prop L Ass'n Q J at 47-49 (cited in note 14).


\bibitem{Kreeger} Karen Young Keeger, \textit{Legal Tussle Over cDNA Libraries May Stall Gene Sequence Efforts}, Scientist 1 (Oct 2, 1995); Karen Young Keeger, \textit{Influential Consortium's cDNA Clones Praised As Genome Research Time-Saver}, Scientist 1 (May 15, 1995).

\bibitem{Bishop} Jerry E. Bishop, \textit{Merck Discloses Details of Plan to Put Data About Genes Into Public Domain}, Wall St J B8 (Sept 29, 1994).

\bibitem{Adler2} Adler, 257 Science at 910-11 (cited in note 14).

\bibitem{Seybold} Seybold, Chief Executive at 18 (cited in note 13).

\end{thebibliography}
the ESTs as probes, as well as smaller portions of those full-length cDNAs that might not even include the disclosed ESTs.\textsuperscript{23} More recently, private owners of cDNA sequence databases have conditioned access to the databases upon agreement in advance to offer either a license or a right of first refusal to any resulting intellectual property rights.\textsuperscript{24} These efforts to claim rights to the future discoveries of others raise issues about the fairness and efficiency of the law of intellectual property in allocating rewards and incentives along the path of cumulative innovation. These concerns are particularly compelling to research scientists, who have more than just commercial interests at stake in inventorship disputes.\textsuperscript{25}

Second, there is a counterintuitive alignment of interests in the debate.\textsuperscript{26} It was a public institution, NIH, that initially took an aggressive position in favor of patenting discoveries that some representatives of industry thought should remain unpatented, and it was a private firm, Merck \& Co., that ultimately took upon itself the quasi-governmental function of sponsoring a university-based effort to place comparable information in the public domain.\textsuperscript{27} These topsy-turvy positions in the public and private sectors raise intriguing questions about the proper roles of government and industry in genomics research and about who stands to benefit (and who stands to lose) from the private appropriation of genomic information.\textsuperscript{28}

\textbf{II. Intellectual Property Rights in Public and Private Research}

Research scientists who work in public institutions are often troubled by the concept of intellectual property because it is counterintuitive to them to promote progress in research through exclusive rights in prior discoveries. Their norms tell them that science will advance most rapidly if subsequent researchers enjoy free access to prior knowledge. By contrast, the working assumption behind intellectual property law is that, without exclusive rights, no one will be willing to invest in research and development.\textsuperscript{29}

\begin{itemize}
  \item \textsuperscript{23} United States patent application \#07/716,831 (cited in note 8).
  \item \textsuperscript{24} See Marshall, 266 Science at 25 (cited in note 17); Marshall, 266 Science at 208 (cited in note 12); Marshall, 266 Science at 1800 (cited in note 17). See also The Institute for Genomic Research, \textit{TIGR Human cDNA Database Agreements} (1994) (on file with author); The Institute for Genomic Research, \textit{Human cDNA Database User Manual} (Beta Test) (1994) (on file with author); Human Genome Sciences, \textit{Material Transfer Agreement} (1994) (on file with author).
  \item \textsuperscript{25} See Jon Cohen, \textit{The Culture of Credit}, 268 Science 1706 (1995).
  \item \textsuperscript{27} Bishop, \textit{Wall St J} at B8 (cited in note 19).
  \item \textsuperscript{28} David Dickson, \textit{`Gene map' plan highlights dispute over public vs private interest}, 371 Nature 365 (1994).
  \item \textsuperscript{29} See Rebecca S. Eisenberg, \textit{Proprietary Rights and the Norms of Science in Biotechnology Research}, 97 Yale L J 177, 196-97 (1987).
\end{itemize}
In a commercial setting a standard argument for recognizing intellectual property is that inventions and discoveries are costly to make in the first place, but cheap and easy to copy once someone else has made them. In order to encourage firms to invest in research and development, it is necessary to give them some means of preventing competitors from reaping the benefits of their investment without sharing in the initial risk and cost.

One way of excluding competitors is to keep inventions secret. But secrecy is not always feasible and may be socially undesirable. The patent system provides an alternative strategy for protecting inventions without secrecy. A patent gives an inventor the right to exclude others from making, using, and selling the invention for a limited term, twenty years from the application filing date in most of the world. But, in order to get a patent, the inventor must make a full disclosure of the invention that is adequate to enable others to make and use it.

Within the realm of industrial research, it is plausible that the patent system promotes greater investment in research and development and greater disclosure of research results than would otherwise occur. This is less clear in the case of public sector research, which presumably need not be profitable to be funded, and which, in the absence of patent rights, would generally be freely disclosed rather than kept secret.

The argument for patenting public sector research discoveries is a variation on the standard justification for patents in the commercial setting, with an emphasis on the post-invention costs and risks involved in taking a new discovery out of the laboratory and developing it into a successful commercial product rather than the pre-invention costs of making the discovery in the first place. The argument is that post-invention development costs typically far exceed pre-invention research outlays and that firms would be unwilling to make this substantial investment in product development without some assurance of protection from competition if the product proves to be successful. Patents thus facilitate the transfer of new technology to the private sector by providing exclusive rights to preserve the profit incentives of innovating firms.

32. See Eisenberg, 5 Risk: Health, Safety & Environment at 164 (cited in note 26); Eisenberg, 55 U Pitt L Rev at 633 (cited in note 26).
III. Public and Private cDNA Sequencing

This argument was advanced by NIH while it was pursuing patent rights in the first few thousand ESTs identified by Dr. Venter and his colleagues at NIH. Yet the response of the intended beneficiaries of the NIH patents—the United States biotechnology and pharmaceutical industries—was less than enthusiastic, suggesting that there may be some limits to the logic of promoting private appropriation of the results of publicly-supported research. Perhaps this was an example of the sort of research discovery that might be more effectively exploited, even by industry, if left in the public domain. Ultimately those particular sequences did enter the public domain, after the United States Patent and Trademark Office rejected the NIH patent applications and new leadership of NIH decided not to appeal the rejection.

NIH did not continue to support the massive cDNA sequencing effort that Venter and his colleagues had begun. Instead, the private sector took over the enterprise of cDNA sequencing. In many respects this represented a step through the looking glass for technology transfer in the Human Genome Project and, as the situation unfolds, it becomes curiouseer and curiouseer.

IV. Non-Patent Strategies for Commercial Exploitation of Sequence Databases

Although the owners of these databases are actively seeking patent protection on their sequences and have obtained a few patents on sequences encoding identified peptides with disclosed function, it remains to be seen what, if any, patent rights they will ultimately obtain in sequences for which they cannot yet provide such a disclosure. Meanwhile, the owners have been able to exploit the databases commercially by controlling access to them, in effect using contracts and trade secrecy to protect their intellectual property.

The viability of contract and trade secrecy as strategies for protecting the sequence information has been limited by Merck's strategy of placing similar information in the public domain. Other things being equal, one would expect

35. See Eisenberg, 55 U Pitt L Rev at 641-43 (cited in note 26).
37. See, for example, Human Genome Sciences, Inc., 1994 Annual Report (“As of April 1, 1995, HGS had filed patent applications with the United States Patent and Trademark office covering over 70 full-length genes and substantially all of the partial gene sequences the Company has discovered. All of these patents are still pending.”)
38. Human Genome Sciences has obtained four such patents to date. United States patent #5,502,969 (Mar 26, 1996); United States patent #5,504,003 (Apr 2, 1996); United States patent #5,506,133 (Apr 9, 1996); United States patent # 5,556,767 (Sept 17, 1996).
39. For an analysis of the patentability of cDNA fragments with no known function, see Eisenberg and Merges, 23 Amer Intell Prop L Ass'n Q J 1 (cited in note 14).
the commercial value of the private databases to decline as the information in the public domain increases. Trade secrecy ends when information becomes publicly available. Moreover, licensees may hesitate to sign restrictive agreements to obtain access to a database if it appears that comparable information might soon be obtained elsewhere on an unrestricted basis.

However, there are differences in the information available from the public and private sources, differences that evidently leave the private database owners with something to sell. Although one might expect that the window of opportunity to sell access to the private databases would close as comparable information becomes available in the public domain,\(^4^0\) commercial firms have continued to sign up for access to the private databases even as the public domain database has grown.\(^4^1\)

Despite the growth of the public database, the private databases remain significantly larger.\(^4^2\) Inasmuch as all the information that enters the public database promptly becomes available in the private databases as well,\(^4^3\) the public database can never contain more information than the private databases. The private database owners also claim to offer superior products in that they have assembled contiguous fragments into longer sequences, they provide more complete annotations for the sequences, including information about expression in different types of tissue, they provide sequence information from customized cDNA libraries derived from tissue types of interest to their subscribers, and their sequence information comes with high-powered bioinformatics capabilities and user-friendly software. Ironically, Merck’s investment in enhancing the public database may have enhanced the value of the private databases as a resource for discovery, not only by contributing further data to make the information in the private databases more complete,\(^4^4\) but also by creating a


\(^{42}\) Elyse Tanouye, *Gene Pioneer Opens His Databank . . .*, Wall St J B1, B16 (Sept 28, 1995).


\(^{44}\) In addition to augmenting the volume of available sequence information, the free availability of the sequence information in the public database has enabled the research community to contribute new information about the sequences, which the private database owners may also freely incorporate into their own databases. For example, sequences in the public database are being mapped to specific locations on chromosomes by a consortium of researchers, Mark S. Boguski and Gregory D. Schuler, *ESTablishing a human transcript map*, 10 Nature Genetics 369 (1995), and the mapping information is also made promptly available in the public domain, enhancing the value of the database to researchers who seek to identify disease genes through positional cloning techniques. Positional cloning refers to techniques for locating a gene responsible for disease on the basis of its map position within the genome as determined through linkage analysis of the DNA of members of families in which the disease is inherited. See Francis S. Collins, *Positional
deluge of information that enhances the value of the complementary proprietary bioinformatics capabilities that the private database owners offer to their clients.45

The value of the public database could be limited by the pending patent applications of the private database owners. If these applications ripen into issued patents, they could subsequently preempt the use of any sequences that they cover, even if those sequences are publicly disclosed prior to issuance of the patents, so long as the patent applicants are able to establish their priority. Because United States patent applications are maintained in confidence until a patent is issued, it is impossible to determine at this stage what sequences have been the subject of patent applications. There are also significant legal questions about which if any of the sequences will qualify for patent protection.47 Those who make use of sequences that are placed in the public database today thus risk facing a future injunction if those sequences turn out to be patented by HGS or Incyte on the basis of previously filed patent applications. Of course, the same uncertainty applies to sequences obtained from the private databases—a sequence obtained by an Incyte database subscriber might turn out to be covered by a previously filed HGS patent, for example. But because the Merck initiative got off to a late start, its sequences are particularly likely to be covered by prior patent applications from the other firms.

V. Exclusive Licensing, Nonexclusive Licensing, and the Public Domain

Meanwhile, in the absence of issued patents, the private database owners may be able to leverage their control over access to the databases today into a valuable proprietary position in subsequent research discoveries tomorrow. The actions of HGS, Incyte, and Merck show three distinct models of how to exploit unpatented information: exclusive licensing, nonexclusive licensing, and dedication to the public domain. Since each of these approaches comes out of the private sector, we can assume that each firm believes, rightly or wrongly, that its strategy will maximize the value it obtains from the information. The strategies are quite different, yet they are interdependent, and it is still too early to tell how each will pay off. But we can see where different firms are placing their bets and we also have some idea of the size of those bets.


46. 35 USC § 122 (1994).

47. See Eisenberg and Merges, 23 Amer Intell Prop L Assn Q J at 3-20 (cited in note 14).
A. Exclusive Licensing

The HGS strategy, at least as initially pursued, approximates an exclusive licensing model. In 1993, HGS entered into a Collaboration Agreement with SmithKline Beecham ("SB")\(^{48}\) giving SB exclusive rights to develop and market protein therapeutic products and diagnostic products out of the information in the database, plus an equity position in HGS, in exchange for payments totaling $125 million, plus royalties on product sales.\(^{49}\) The agreement did not cover gene therapy or anti-sense products; HGS has entered into separate collaborative agreements with other research partners for the development of products in these areas.\(^{50}\) Moreover, during the period of SB’s exclusive license, outside investigators have been able to obtain access to the HGS database by signing a "Material Transfer Agreement" granting "a sole and exclusive worldwide right and license" to HGS to develop any resulting products on terms to be negotiated in the future.\(^{51}\)

An obvious advantage of this exclusive licensing strategy, at least from the perspective of HGS, is that it generated considerable revenue—SB placed what looked at the time to be a very large bet. An obvious concern with the exclusive licensing approach is that restricting access to the database to such a degree could limit the amount of value that is extracted from the information during the term of the exclusive license. There is only so much information that a single firm and its small circle of collaborators can use or even evaluate in a finite time period.

More recently, HGS and SB have departed from their original exclusive licensing strategy in favor of allowing more firms to tap into the database.\(^{52}\) Within the past year, HGS and SB have entered into collaborative agreements to allow four additional pharmaceutical firms—Takeda Chemical Industries,\(^{53}\) Merck KGaA (not related to Merck & Co.),\(^{54}\) Schering Plough\(^{55}\) and

---

48. A version of this agreement with certain confidential terms omitted was placed on file with the Securities and Exchange Commission as an exhibit to the prospectus for HGS's initial public offering of common stock in 1993 and is on file with the author.
51. See, for example, Human Genome Sciences and the University of Michigan, Material Transfer Agreement at ¶ 14 (Sept 14, 1994) (on file with author).
53. See Takeda to access gene information, Biotechnology Newswatch 13 (Sept 18, 1995).
Synthelabo SA\(^5\) to share access to the database, evaluate the information, and develop related products. These new collaborators will make total payments of at least $140 million to HGS and SB,\(^7\) not including milestone payments and royalties. In addition to bringing in new revenue, it is likely that these new agreements will increase the amount that is learned from the database by expanding the universe of investigators who are able to make queries.

Some of the sequences in the HGS database were obtained by its nonprofit affiliate, TIGR, under the terms of agreements giving HGS intellectual property rights in TIGR discoveries. These agreements permitted TIGR to publish its DNA sequences after specified time periods following disclosure to HGS. Pursuant to these retained rights, TIGR established a separate database, TIGR Human cDNA Database ("HCD"), to be made available to investigators in academic and non-profit institutions under "Database Access Agreements" with less restrictive terms than those provided in the HGS Material Transfer Agreement.\(^8\) The TIGR HCD initially included most of the TIGR-generated sequences along with those HGS sequences that corresponded to sequences that had already been disclosed by others in a public database. The terms of access for unpublished sequences are more restricted than for published sequences, requiring that the investigator's institution first sign a "Database Option Agreement" granting HGS a sole and exclusive option to obtain a sole and exclusive or a nonexclusive worldwide royalty bearing license to resulting products.\(^9\) Sequences that have been disclosed or partially disclosed in a public database are available on less restricted terms,\(^6\) but initially even for these sequences investigators were limited to a specified number of queries that could be recorded, stored, and monitored by the Database Manager. No outside investigators could trawl through the database or manipulate its contents at will and commercial investigators could not obtain access to the database at all. As the information available in public databases has expanded, a larger portion of the TIGR HCD has become available on less restricted terms and TIGR has recently announced its intention to make the entire database freely available on the World Wide Web as

---


57. Schering-Plough agreed to pay $55 million, PR Newswire, *Schering-Plough* (cited in note 55), Merck KGaA agreed to pay $50 million, PR Newswire, *Human Genome Sciences* (cited in note 54), and Synthelabo agreed to pay $35 million, PR Newswire, *SmithKline Beecham* (cited in note 56). The payments called for in the agreement with Takeda were not disclosed.

58. The terms of access are described by TIGR on the World Wide Web at http://www.tigr.org/db/hcd.html#new ("TIGR Web Page").


60. Institute for Genomic Research, TIGR Human cDNA Database Agreements: Level I Access Agreement § 3 (1994); Institute for Genomic Research, Human cDNA Database User Manual (Beta Test) at § 3.2 (cited in note 24).
of April 17, 1997.

B. NONEXCLUSIVE LICENSING

Incyte pursued a nonexclusive licensing approach for access to its proprietary LifeSeq™ database from the outset, offering database subscription agreements to as many firms as will take them, at a considerably cheaper price than the $125 million SB paid for its deal with HGS. So far eleven firms have signed up in agreements with varying terms. Pfizer and Upjohn each signed agreements involving payments in the range of $20 million to $25 million, including amounts paid for the purchase of Incyte stock, plus royalties on product sales. Novo Nordisk and Hoechst have signed similar agreements on undisclosed financial terms. In December 1995, Incyte announced that it had signed a database agreement with Abbott Laboratories that it characterized in a press release as "the largest financial commitment by a subscriber to date," although again the specific financial terms were not disclosed. When Incyte announced its sixth subscriber, Johnson & Johnson, in January 1996, the accompanying press release claimed that the six subscribers will provide in the aggregate a minimum of $100 million toward the development of Incyte's proprietary genomic databases, excluding contingent payments such as milestones and product royalties. Since that announcement, Incyte has entered into additional agreements with Hoffmann-La Roche, Zeneca, BASF AG, Schering AG, Berlin, and Eli Lilly. Incyte's subscribers have each placed smaller individual bets than SB did,

---

61. See TIGR Web Page (cited in note 58).
62. Comparison of the terms of the different agreements is complicated by the fact that some of the agreements involve the creation of satellite databases of sequences expressed in different tissue samples of particular interest to the subscriber, and some involve access to other database products, including the LifeSeq™ database of full-length genes and PathoSeq™ database of microbial DNA sequences.
63. Incyte Pharmaceuticals, Inc., 1994 Form 10-K.
64. Incyte Signs Database Deal With Novo Nordisk, Market Letter (Sept 4, 1995);
65. Incyte Pharmaceuticals, Inc., Press Release: Incyte and Abbott Sign Genomic Database Subscription Agreement: Agreement to Include Microbial Genomes (Jan 2, 1996). The agreement with Abbott Laboratories includes payment for sequencing microbial genomes, which may explain why it represents the largest financial commitment by a subscriber to date.
71. Incyte Pharmaceuticals, Inc., Press Release: Incyte and Lilly Enter Into Genomic
but there are considerably more of them, and it is quite possible that Incyte's nonexclusive licensing strategy has already generated more revenue for the database owner than HGS's exclusive licensing strategy. One might have expected Incyte's window of opportunity to sign up new subscribers to close as the Merck-sponsored sequencing effort at Washington University expanded the competing public database, yet all but the first of Incyte's subscribers signed up after Merck announced its competing effort. The most recent subscriber signed up even as the Merck-sponsored effort approached completion.

From a broader social standpoint the more interesting question is not the size of the bets but the ultimate payoffs. Which approach will yield more discoveries or more commercial products? Although the nonexclusive strategy seems more likely to take advantage of the different capabilities of different firms, a drawback of the Incyte approach is that the company has not yet figured out a way to make its database directly available to academic investigators without undermining its value to Incyte's corporate subscribers. Academic investigators may only obtain access to the Incyte database by collaborating with a corporate database subscriber.

C. PUBLIC DOMAIN

The Merck strategy of putting cDNA sequence information in the public domain is, at least at first glance, the most puzzling from a strategic standpoint. Rather than creating its own proprietary database or subscribing to another firm's database, Merck has chosen to finance a human cDNA sequencing project at the Genome Sequencing Center at Washington University to create the Merck Gene Index as a public resource. The results of this sequencing effort are deposited daily in databases that are freely accessible to the public, without any delays in disclosure to secure patent rights or even advance access by Merck's own researchers. It is easy to justify such a move rhetorically as a public-spirited contribution to the larger research enterprise, but it is nonetheless worth considering how this strategy might advance Merck's own commercial interests.

One advantage to Merck of putting the information in the public domain is that it allows it to generate the information more cheaply—indeed, almost unbelievably cheaply. Merck has placed a relatively small bet, reportedly some-

Database Partnership (Dec 18, 1996).
72. See notes 40-41 and accompanying text.
73. See Merck & Co., Inc., Press Release: First Installment of Merck Gene Index Data Released to Public Databases; Cooperative Effort Promises to Speed Scientific Understanding of the Human Genome (Feb 10, 1995).
74. Id at 2.
75. See, for example, Merck & Co., Inc., Press Release: Merck to Make Comprehensive Database of Genetic Information Publicly Available (Sept 28, 1994) ("The Merck Gene Index is the latest example of Merck's century-long commitment to providing comprehensive and authoritative scientific and medical information as a public service on a not-for-profit basis.")
where under $10 million. This is a fraction of the amounts spent by non-exclusive subscribers for access to Incyte's database, and a tiny fraction of the amount spent by SB for its exclusive deal with HGS. By positioning itself as a public benefactor, Merck is able to make use of the capabilities of the Genome Sequencing Center at Washington University, in effect benefiting from the experience and facilities of a group that has received considerable public funding. Merck has also been able to obtain access at nominal cost to research materials that, in all likelihood, it would have had to pay a premium for if it were trying to assemble a private database. The cheap price tag for generating the information also suggests that Merck may not be giving up that much from a competitive standpoint by foregoing the opportunity to retain it as a proprietary resource; Merck's competitors could presumably create a comparable resource within the same price range.

Apart from generating the sequence information more cheaply, Merck claims that it expects to derive more benefit from the information by distributing it widely. As Merck representatives explain, the sequence information will not yield products for commercial development until further fundamental biological research is done to understand the functions and biological pathways associated with the partially sequenced genes. Merck's comparative advantage does not lie in performing this fundamental research, but rather in developing specific drugs at a later stage in the research and development process. By promptly placing the sequence information in the public domain, and thereby making it widely available to academic researchers, Merck anticipates benefiting in the long run from the fundamental research of those who use the database. Nothing obligates these researchers to bring any potential products to Merck for commercial development, but Merck is confident that its capabilities and resources will allow it to capture an adequate share of resulting products to justify its modest investment in generating the database. In other words, Merck expects to profit more in absolute terms by making the database publicly available, even if other firms also profit as a result.

Some observers have suggested a more cynical possibility—that Merck seeks to undermine the value of investments already made in existing sequence databases by its commercial competitors. Putting the information in the public domain leaves HGS and Incyte (and their collaborators) dependent on patent rights to protect their proprietary positions in the long run, and Merck may be

77. The accomplishments of the Genome Sequencing Center are summarized on the World Wide Web at http://genome.wustl.edu/gsc/gschmpg.html.
80. The discussion in the text is based in part on telephone conversations with Bennett Shapiro (Oct 17, 1994) and Alan Williamson (Nov 11, 1994).
betting that they will not obtain much in the way of patent rights. Merck's own strategy for making money does not rely on maintaining a proprietary position in cDNA sequences and therefore it has little to lose, and possibly something to gain, by putting such sequences in the public domain. Merck does not have any therapeutic protein or DNA diagnostic products that might require proprietary rights in DNA sequences to be commercially viable. Its proprietary rights in small molecules that it hopes to sell as pharmaceutical products are far more significant to its commercial position than proprietary rights in DNA sequences. This distinction highlights the very different ways in which the sequence information is valued by Merck on one hand and HGS and Incyte on the other. From Merck's perspective, cDNA sequences are research tools for use in drug discovery, not products for sale to consumers. For HGS and Incyte, cDNA sequences are themselves a product in that they supply an immediate source of revenue.

As the Merck Gene Index neared its scheduled completion date at the end of 1996, it did not in fact appear to have undermined the commercial value of the HGS and Incyte databases. Quite the contrary, since Merck announced its plan, Incyte has signed up ten new subscribers and HGS and SB have entered into collaborations with four additional firms. Obviously, the private database owners still have something valuable to offer Merck's competitors.

Apart from its impact on the profit expectations of private database owners and their collaborators, the information in the Merck Gene Index may have considerable social value if other firms and publicly-funded researchers put the information to use. Will the nonproprietary character of the information lead commercial firms to shun the data for fear of being unable to exclude competitors from the market for any resulting products or will the public database be actively and widely exploited? Accessions to the public database GenBank have shown a dramatic increase since the Merck data started coming on line. A big part of the increase has come in the form of anonymous file transfer protocol ("FTP") downloads of the entire database, a form of query that is likely to be popular with commercial users who do not want to risk tipping their hands to competitors by leaving an electronic record of what it is they are looking for.

VI. The Public Domain and the Public-Private Divide

These three different approaches highlight striking differences in the interests of different firms in proprietary rights in the human genome. HGS and Incyte may benefit from a strategy that promotes the private appropriation of DNA sequences; Merck may benefit from a strategy that puts these sequences in the public domain: One firm's research tool may be another firm's end product. Yet in an important sense, the fact that Merck chooses to put cDNA sequences in the

82. Craig, 7 BioWorld Today at 197 (cited in note 41).
83. See notes 53-57 and 63-71 and accompanying text.
85. Telephone conversation with Mark Boguski at the National Center for Biotechnology Information (Aug 1, 1995).
public domain is more instructive than the fact that HGS and Incyte choose to appropriate them as private property. Whenever new property rights come into view, someone will step forward to claim them, and it is unsurprising to hear the claimants tell a story about how private ownership will enhance the public welfare. It is more uncommon for a private firm to disclaim proprietary rights in the information it generates and to sing the praises of the public domain.

The Merck initiative raises fundamental questions about the boundaries between public and private in research science and product development. In an earlier era, we could have given a coherent account of these boundaries in theory, however blurred they may have been in practice. Public research tended to focus on the pursuit of fundamental knowledge of a character that was not readily appropriable by a private owner, and it was generally believed that such knowledge would have the greatest social value if it was widely distributed with no restrictions on its use. Private research, by contrast, tended to focus on narrower applications of scientific principles that were readily appropriable by innovating firms. These firms required proprietary rights to protect them from competition and to make their investments in research and development profitable. Publicly supported research was presumptively placed in the public domain, while privately supported research was typically appropriated as intellectual property.

Today, these boundaries are more difficult to maintain, particularly in fields of such obvious commercial interest as genetics. Researchers in the public and private sectors have often found themselves working on the same problems, whether competitively or collaboratively, and the prevailing wisdom is that institutions performing publicly-sponsored research should patent their discoveries whenever it is profitable to do so. In this environment we lack a clear story about when it makes sense for the public to sponsor research and when it makes sense to dedicate new knowledge to the public domain.

When public policy promotes private appropriation of research results as intellectual property even when they emerge from public sector research, it is easy to lose sight of the public and private benefits of disseminating information in the public domain. The most obvious of these benefits are that free availability encourages widespread use of information and minimizes transaction costs. Travel on a freeway is both cheaper and faster than travel on a tollway with numerous tollbooths. Similarly, research and development is cheaper and faster if it uses resources that are freely available than it is if the road to product development requires frequent stops to negotiate licenses for access to prior discoveries. Free roadways can also enhance the value of private property by making it more readily accessible. Thus free dissemination of information on the Internet helps firms attract customers. Of particular relevance to research science, a vigorous public domain can supply a meeting place for people, information, and ideas that might not cross paths in the course of more organized, licensed encounters. And finally, of particular relevance to fields of research that draw heavily upon work that is done in the public sector and the academy, information in the public domain is accessible to relatively impecunious users who would otherwise be priced out of the market. Thus, for example, if academic research-
ers are particularly important to the progress of research in a field, as Merck evidently believes they are for understanding the human genome, then the overall research enterprise could be significantly retarded by property rights that restrict their access to essential resources and information.

If ESTs in fact have more social value in the public domain than in the hands of private owners, perhaps government research sponsors should have taken upon themselves the burden of supplying this resource to the public rather than leaving it to the private sector. On the other hand, perhaps the extent of private sector interest in supporting large-scale cDNA sequencing indicates that this work does not require government funding and that public resources would be better spent on other projects. The Merck initiative invites the optimistic conclusion that there are limits to how far the government can go wrong—that if the stakes are high enough, someone in the private sector may find it worthwhile to correct for any errors in judgment on the part of the government and maybe even to pick up the tab.

Yet it would be foolish to conclude on the basis of this extraordinary episode that we can rely on the private sector to create a public domain whenever that is the most efficient way to exploit new information from a broader social perspective. Perhaps a better lesson to draw is that we may have underestimated the value of a rich public domain to the private sector as well as to the public sector, and that we may need to reconsider the limits of private appropriation of new information as a means of promoting commercial development.