

The Ethics of Access to Online Genetic Databases: Private or Public?

by A. Marks, B.A., Karen K. Steinberg, PhD.

Abstract

With the sequencing of the human genome comes the promise of advances in medical science. For this promise to be fully realized, researchers must have access to information resulting from this landmark endeavor as well as from subsequent research initiatives. However, because genomic sequences are potential sources of profit for the biotechnology and pharmaceutical industries, many private companies seek to limit access to this information. Some argue that this will impede scientific progress, while others argue that the privatization of genetic information is needed to assure profits and generate the considerable funding necessary to bring therapeutic products to the market. In this paper, we present arguments for both sides, and conclude that both private funding and public access to information are important in genetic research. Precedents for compromise are necessary, as is increased dialog between private and public interests in order to ensure continued advancements in genetic science and medicine.

The completion of the rough draft of the human genome in March of 2001 was heralded by many as a scientific landmark. After the initial announcement, the popular media was flooded with articles predicting the advances in medical science that would be soon follow.^[1] While few can doubt the eventual benefits that will result from genetic research, practical applications of genetic technologies such as gene therapy lie mostly in the future.^[2] However, that future seems to be moving closer and closer as our knowledge of genetics is increasing faster than ever.

Establishing Genetic Databases: What are the Goals?

Since the 1970s, a concerted effort has been made to make human genome sequence information freely accessible to researchers around the globe, and projects such as the Human Genome Initiative (HGI)^[1] have been created with this express purpose in mind.^[3] From the outset, HGP has emphasized that data obtained from HGP-funded research must be publicly available. The rationale for such endeavors is based on the idea that our ability to expeditiously and effectively increase our knowledge of genetics depends on the ability of researchers to access current information.^[3] However, a subsidiary, but explicit, goal of those responsible for creating and funding the HGP is the creation of technology and economic benefit.^[4] Indeed, in the past decade, growing commercial interest has resulted in the creation and upkeep of private genetic databases. As a result, many members of the scientific community worry about the impact that such databases will have on research ventures.^[5] In this paper, we will provide an overview of the issues involved with access to genetic databases.

The Access Debate

Common thought in the past has held that in order to allow for maximum research potential, we should allow all interested parties free access to genetic databases.^[5] This has been, and remains, the dominant viewpoint in scientific circles. However, this belief has been challenged in the past 10 years, and the question of access has become more complicated.^[6] Although no one doubts the value of public domain databases, there has been a push by commercial interests in the past decade to create private databases to recoup and profit from the investments of the private sector.^[7] Such databases are thus an important tool by which biotechnology companies make an early profit from genetic research. Without the promise of revenue, some argue, genetics and other fields of biotechnology could lose an enormous amount of funding from the private sector, and thus slow down the development of practical applications of genetic research.^[8]

Making the case for public access are the HGP and organizations such as the NIH, Wellcome Trust, and

other research groups who say that access to genetic databases should be public so that genetic information can be disseminated in the most effective fashion. Without public access, such groups argue, scientific research and advancement could be severely stifled.^[9] In March 2000, these concerns prompted then President Clinton and Prime Minister Tony Blair to urge private companies to “make raw genomic data publicly available.”^[10] At the heart of the issue of access to genetic databases lies the larger problem of the commercialization of the human genome.

Private Databases

In the last years of HGP’s attempt to sequence the human genome in its entirety, a great deal of attention was given to the efforts of a competitor, the for-profit company Celera.^[11] Celera, a company started by former NIH employee J. C. Venter, announced that not only would Celera map the human genome but that it would also finish before the HGI, thanks to Celera’s use of a whole-genome shotgun sequencing approach.^[11] Celera also later announced that the sequence data generated by its efforts would be stored in a private Web-based database. Research groups would gain access to this database by paying a fee ranging from \$US5,000 to \$US20,000 annually, depending on whether the group was non-profit or commercial.^[12] The draw of the Celera database is not in the sequence data, which is publically available on Web-based databases. Rather, Celera hopes to offer subscribers the software tools, supercomputing power, and integrated links (such as easy links to research histories, exon/intron structure, and polymorphisms) that the public databases are not able to provide or maintain.^[12,13]

Private genetic databases are not new concepts, and Celera was not the first to market such a product. While the degree of access and the nature of the access agreements vary from database to database, these databases have already generated millions of dollars in revenue.^[5] In 1996, the Maryland biotechnology company Human Genome Sciences (HGS) sold the three-year exclusive right of access to its database of cDNAs to SmithKline Beecham for \$125 million dollars. The deal gave SmithKline Beecham the exclusive right to develop and market therapeutic and diagnostic products using information from the HGS database.^[3] Other firms, such as Incyte Pharmaceuticals, have offered non-exclusive licenses (i.e. other companies can also purchase access to the information) for \$20-\$25 million to companies such as Pfizer and Pharmacia Corporation.^[5] Arrangements like these have shown how private databases can allow investors to generate a great deal of revenue early on in the research and development process, thus making such databases an attractive strategy to many biotechnology companies.

Willingness to Pay for Access

Why are companies such as SmithKline Beecham and Pfizer paying for access to private databases when they could instead support the formation of free, public domain databases and save themselves millions of dollars? Because in the commercial fields of biotechnology and drug development, time is of the essence. As our knowledge of genetics and genetic technology increases, companies are scrambling to patent novel and commercially-useful gene sequences.^[2] By the end of the year 2000, over 25,000 DNA-based patents were issued and we can expect that this number will continue to increase over the next decade as private companies and academic centers both move to secure genetic information.^[4] The first group to discover novel mutations or “useful” genetic sequences (i.e., specific-function sequences, single-nucleotide polymorphisms, or expressed sequence tags) has the right to patent the sequence and reap any financial benefit that the sequence might generate.^[14] Having access to a database that your competitors do not means having the first opportunity of discovering and patenting sequences that may serve as targets for useful diagnostic and therapeutic products. Celera and others have already applied for thousands of patents on genetic sequences discovered during the mapping of the human genome, and companies such as SmithKline Beecham and Incyte Pharmaceuticals hold hundreds of gene patents.^[13] It would appear, then, that the use of private databases is an effective tool by which companies can get an early start in the patent race.

Patenting Genes: Legal Protection of Assets

It is important in our understanding of the significance of private genetic databases to be familiar with the scope and status of genetic patents. Although guidelines of the U.S. Patent and Trademark Office (USPTO) require that DNA and RNA sequences be shown to have a useful application in order to be

patented,^[14] the USPTO upholds the eligibility of the majority of genes for patenting on the basis of law and judicial precedent. In fact, over 9,000 gene-related patents were granted between 1997 and 1999,^[14] and industry, universities, and government agencies are actively seeking to patent nucleotide base sequences. A recent U.S. Supreme Court decision, referred to as *Festo* (for *Festo Corp v. Shoketsu Kinzoku Kogyo Kabushiki Co*), will affect the interpretation of some patent claims, including those related to DNA and RNA sequences.^[16, 17] Briefly, patent law provides two zones of protection. The first is a zone of protection that is well-defined and obvious to the patent holder and competitors. A second zone is the fuzzy area in which a competitor can produce a similar product with only 'insubstantial' changes that may or may not infringe on the patent. Cases that fall in the second zone must be decided by the courts. Courts usually allow competitors to make products with 'substantial' changes made to avoid patent law but not those with 'insubstantial' changes. The degree to which products are judged 'substantial' and 'insubstantial' is determined by the courts. The lower court's *Festo* decision severely restricted the flexibility of the fuzzy zone and the scope of genetic patents and, as a result, was seen as a liability to patent holders. In fact the Supreme Court chose a compromise proposal between the restrictive *Festo* decision and the more liberal interpretation of the fuzzy zone. This decision has the potential to allow researchers more flexibility when working with previously-patented genes while still offering some protection to the patent holder.

From a practical point of view, patenting genes and creating private databases encourage progress in medicine and science by assuring the participation and funding of industry in drug discovery. This participation on the part of industry has already greatly accelerated the pace of genetic research.^[18] In the expensive field of genetics and biotechnology, the cost of developing therapeutic products can be prohibitively expensive, and the investment of private companies has often been critical to developing new technologies. Some biotech companies argue that without the incentive that privatization and patenting offers, therapeutic products of genetic research are much less likely to reach the public.^[8] One study suggested that without the financial incentives that patents and privatization of information offer, 60% of pharmaceutical products would not have been able to reach the market.^[19] Clearly, the private sector plays an important role in bringing practical application of biotech research to the market, and many argue that by encouraging commercial funding, private databases are beneficial for the research initiative.

Public Databases

As mentioned before, programs such as HGI and the Human Genome Organization (HUGO) aim to post genetic information on the internet in publically accessible databases. This information can be accessed by anyone with an internet connection, and new information on genes and genetics is being posted daily. A number of more specific public databases exist, and more are planned (Table 1). Today, databases contain information on genetic diseases, on genes and their location, and on mutations existing on already cloned genes. These databases allow scientists, at little or no cost, to follow current genetic research, compare newly-sequenced genes with existing data, and quickly and efficiently post the results of their own studies.^[19]

Since the 1970s, scientists and politicians alike have agreed that the results of the efforts to map and sequence the human genome should be made freely available to the public.^[3] To further the goal of keeping data freely accessible, public domain genetic databases have existed for decades, starting with the Human Gene Mapping (HGM) library.^[3] Since then, biology and the internet have enjoyed a happy and very productive partnership, and almost every major genetic database now exists in an online form.^[21] Currently, there exist several hundred public genetic databases, and in addition to genetic sequence databases, there are databases for gene expression data, population polymorphism data, and biochemical pathway data.^[21, 22] These and other types of data are commonly used in various types of genetic and biochemical studies and are important in the effort to discover and characterize new genes.^[21] Groups such as the National Center for Genome Resources (NCGR) and the European Molecular Biology Network (EMBnet) are constantly striving to find more efficient ways to integrate different databases, to allow researchers easy access to a wide array of information.^[23]

Cost Concerns

Because genetic databases are important tools with which researchers can perform homology studies and other sequence comparisons, some scientists worry that the creation of private databases will keep key information from the scientific community, thereby severely limiting research potential. Furthermore, it is feared that the actions taken by private industry, such as patenting of nucleotide sequences and creating private databases, will greatly increase the cost of basic research. This rise in cost could further limit access for many of the relatively poorly funded research groups (e.g. researchers new to a field or research in areas of relatively limited interest). In a field such as genetics, which depends heavily on research in the public sector, the reduction of public research participation could limit the number of research avenues being pursued, severely retarding the overall research endeavor.^[5]

The Conflict of Public versus Commercial Interests

The debate, then, can be understood as primarily involving different conceptions of how to best foster the research endeavor in genetics. Those in support of private databases believe that genetic research needs the funding from commercial interests, thus the privatization of data is in the best interests of all those who hope to reap the benefits of genetic technology. Those against private genetic databases believe that the interests of science will best be served by enabling all groups to have free access to the information, thus maximizing the number of groups able to participate in genetic research.

In support of their view, commercial interests are often quick to point out that past situations involving the privatization of scientific information have rarely led to restrictions in the public sector.^[24] In the 1960s, similar fears were expressed over the patenting of certain basic polymers. In fact, the commercial patenting of polymers such as aliphatic mono-olefins did not result in the hindering of polymer research.^[24] In another example, over 1,000 HIV virus genes are currently patented, and the research in this area has not been stifled.^[15] Furthermore, industry holds that the majority of the patent disputes that have occurred in the fields of biotechnology have been between commercial competitors, and academic researchers have been largely left alone.^[9] These facts, some argue, suggest that steps to privatize information, such as the creation of private databases, would not be overly damaging to public research.

However, there is evidence to the contrary. We have already seen instances of private companies barring research on the basis of patent infringement, so many scientists are understandably worried that similar conflicts will arise over the use of private databases.^[15] The well-publicized patent infringement case concerning Myriad Genetics and certain research groups involving Myriad's *BRCA1* patents is of particular concern. Myriad's patents on *BRCA1* prohibited research groups from being reimbursed for performing *BRCA1* testing, and several research protocols have been halted because of this limitation.^[15] Another case involves Glaxo SmithKline, which owns patents on the *ApoE4* allele, an allele thought to put carriers at an increased risk for Alzheimer's Disease. Testing for the *ApoE4* allele can identify increased risk of developing Alzheimer's Disease, with a positive predictive value of 94-97%.^[25, 26] Glaxo SmithKline has patented a test for the *ApoE4* allele, and several academic genetics labs have been told to stop offering their own version of the *ApoE4* test for a fee.^[15] Since we have already seen conflicts on the basis of patented information, it is not difficult to believe that private companies would strictly enforce the private nature of their databases to the exclusion of public research initiatives.

It is important to note that some biotech firms are actually in favor of public databases.^[15] The major pharmaceutical firm, Merck Research Laboratories, assumed the sponsorship of a university-based effort to place in the public domain genetic sequence information comparable to that found in private databases.^[4] According to Merck, their reasoning behind such a move is that the majority of genetic information will not yield products for commercial development until further research is done. Therefore, it is in Merck's best interest to allow public access to sequence information so that more research can be done, and then Merck can develop specific drugs later in the research and development process.^[5] In addition, a number of large private companies have joined with the Wellcome Trust program to form The SNP Consortium LTD, a group that plans to spend \$45 million over the next two years to find and patent hundreds of thousands of single nucleotide polymorphisms (SNPs) and to create a public database of this information.^[27] The actions of Merck and the companies collaborating with Wellcome Trust indicate the willingness of some industry leaders to ensure that genetic information remains in the public domain and acknowledge that genetic information best serves the research interest by being publically available.

Conclusion

It may seem at this point that the debate over access to genetic databases is almost moot. Patent law and economic realities ensure that private databases will continue to flourish, and so long as genetics and related technologies hold the promise of products and profit, there will be those interested in marketing the results. However, what has yet to be determined is the degree to which private databases control genetic information. Already, the Wellcome Trust has formulated policy that prohibits research groups from using Wellcome Trust funding to purchase access rights to Celera's database,^[28] and a recent survey revealed that over 75% of human geneticists are against the privatization of genetic information.^[29] Groups such as The SNP Consortium LTD are taking aggressive action to protect certain information from being privatized, and other companies have acknowledged the importance of keeping genetic databases open to the public.^[3] Such actions reveal the strong feelings of the research community in favor of publically accessible data and against the further privatization of genetics.

Of course, there are other concerns at stake in the conflict over genetic databases, which is only one facet of the larger issues of the commercialization of the human genome. Since 1994, private funding in genetic research has exceeded federal funding, and many worry about the influence such a dominant presence will have on the direction of research.^[18] It is feared, even expected, that as commercial interests play a larger role in genetics, more and more genetic information will become patented. Private databases can be understood as a way in which the human genome is becoming further commercialized, the means by which genetic information is used to generate revenue for companies who have made an investment. The concerns raised by opponents of private databases are reiterations of the worries many have over the general privatization of genomic information.^[30,31]

The creation of private genetic databases is explicitly aimed at allowing pharmaceutical and biotechnology companies to take out patents on useful gene sequences. As we have seen, patents on genetic sequences and gene tests have already impinged on academic research, and it can be expected that as more genetic information is held by private interests, the tension between private and public sector groups will increase. However, we have also seen that such tension does not necessarily have to be the result of private sector involvement in genetic research. We appreciate the role of private funding in the development of therapeutic products from genetic research. Without the involvement of private companies, important therapeutic products may not reach the market. At the same time, private companies may not be able to develop these therapeutic products without the initial basic research provided by public research organizations. Increased collaborative efforts, such as that of the Wellcome Trust, are needed between private and public research groups to ensure continued advancements in genetic science and medicine. We hope that as the patent rush continues and as our knowledge of genetics and related technologies increases that both private and public interests will understand the reality that genetic research as a whole is best served by a collaboration between both groups.

References

1. Semple, CA. Bases and Spaces: Resources on the Web for Accessing the Draft Human Genome- After Publication of the Draft. *Genome Biology* 2001 June; 2(6): 1-7
2. The scramble to patent human genes. *Nature Neuroscience* 1999 Sept; 2: 773
3. Pearson PL. Genome mapping databases: data acquisition, storage and access. *Current Science* 1991; 1: 119-123
4. Cook-Deegan RM, McCormack SJ. Intellectual property: patents, secrecy and DNA. *Science* 2001 July; 293 (5528): 217
5. Eisenberg R. Intellectual property issues in genomics. *Bioinformatics* 1996 August; 14: 302-307
6. Knoppers BM Laberge CM. Ethical guideposts for allelic variation databases. *Human Mutation* 2000; 15: 30-35
7. Goodman L. Unlimited access- limitless success. *Genome Research* 2001; 11: 637-638
8. Lytle J. Issues concerning ethical conduct and genetic mapping raised at Montreal meeting. *Canadian Medical Association Journal* 1997 February; 156 (3): 411-412
9. Butler D and Smaglick P. Celera genome licensing terms spark concerns over 'monopoly.' *Nature* 2000 January; 403: 231

10. Katzman S. What is patent worthy? European Molecular Biology Organization Reports 2001; 2 (2): 88-90
11. Pennisi E. Academic sequencers challenge Celera in a sprint to the finish. Science 1999 March; 283: 1822-1823
12. Butler D. US/UK statement on genome data prompts debate on 'free access.' Nature 2000 March; 404: 324-325
13. Celera Discover System Overview, <http://www.celera.com/genomics/academic>
14. Bonetta L. Raising the bar for genetic patents. Current Biology 2001; 11 (2): R115-R116
15. Harris RF Patenting genes: is it necessary and is it evil? Current Biology, 2000; 10 (5): R174-R175
16. Williams KM. New draconian restrictions on U.S. patent scope: FESTO. Palmer & Dodge, LLP 2001. (<http://www.palmerdodge.com/dspSingleArticle.cfm?Articleid=352>)
17. Williams KM. Genome success means difficult patent questions. Palmer & Dodge, LLP 2000. (<http://www.palmerdodge.com/dspSingleArticle.cfm?Articleid=175>)
18. Burris JR, Cook-Deegan M, Alberts B. The Human Genome Project after a decade: policy issues. Nature Genetics 1998 December; 20: 333-335
19. Ayme S. Bridging the gap between molecular genetics and metabolic medicine: access to genetic information. European Journal of Pediatrics 2000; 159 (S183): 183-185
20. Adler RG. Genome research: fulfilling the public's expectations for knowledge and commercialization. Science 1992 August; 257: 908-914
21. Skupski MP, Booker M, Farmer A, et. al. The Genome Sequence Database: towards an integrated functional genomics resource. Nucleic Acids Research 1998 November; 27 (1): 35-38
22. Gu Z, Hiller L. Single nucleotide polymorphism hunting in cyberspace. Human Mutation 1998 May; 12: 221-225
23. Norman F. Genetic information resources: A new field for medical librarians. Health Libraries Review 1999; 16: 15-28
24. Doll JJ. The patenting of DNA. Science 1998 May; 280: 689-690
25. Marshall E. The battle over BRCA1 goes to court; BRCA2 may be next. Science 1997 December; 278: 1874
26. Roses, AD. Genetic Testing for Alzheimer Disease: Practical and Ethical Issues. Archives of Neurology 1997 October; 54 (10): 1226-1229
27. Lytton M. Patent problems in the gene pool. Palmer & Dodge, LLP 1999. (<http://www.palmerdodge.com/dspSingleArticle.cfm?Articleid=114>)
28. Stephens D. Wellcome Trust discourages Celera subscribers. Trends in Cell Biology 2001 July; 11 (7): 284
29. Rabino I. How human geneticists in U.S. view commercialization of the Human Genome Project. Nature Genetics 2001 September; 29: 15-16
30. Caulfield T. The commercialization of human genetics: profits and problems. Molecular Medicine Today 1998 April; 4: 148-150
31. Chadwick R, Berg K. Solidarity and equity: new ethical frameworks for genetic databases. Nature Reviews 2001 April; 2: 318-321

Table 1: Popular public-domain genetic databases;* Sequence and allelic variation databases:

GenAtlas:	http://bisance.citi2.fr/GENATLAS/
Human Genome Database:	http://www.gdb.org/
Human Genome project Working Draft:	http://genome.ucsc.edu/
HUGO mutation database:	http://ariel.ucs.unimelb.edu.au:80/~cotton/mdi.htm
OMIM:	http://www3.ncbi.nlm.nih.gov/omim/
International Nucleotide Sequencing Database Collaboration:	http://www.ebi.ac.uk/embl/
Sequence Tag Alignment and Consensus Knowledge Database:	http://www.sanbi.ac.za/Dbases.html
GenLink:	http://www.genlink.wustl.edu/
GenBank (BLAST tools):	http://www.ncbi.nlm.nih.gov/Genbank/
TIGR Human BAC Ends:	http://www.tigr.org/tdb/humgen/humgen.html
Breast Cancer Gene Database:	http://condor.bcm.tmc.edu/ermb/bcgd/bcgd.html
MCA/MRS:	http://www.nlm.nih.gov/mesh/jablonski/syndrome/
Pharmacogenetics Knowledge Base (PharmGKB):	http://www.pharmGKB.org/

*There currently exist so many genetic databases that a databases of genetic databases has been created, the Public Catalogue of Databases. Go to <http://www.infobiogen.fr/services/dbcat/> for a more detailed list of specific genetic databases.

BAC = bacterial artificial chromosomes; BLAST = Basic Local Alignment Search Tool; HUGO- Human Genome Organisation; MCA/MRS = multiple congenital abnormality/mental retardation syndromes; OMIM = Online Mendelian Inheritance in Man; TIGR = The Institute for Genomic Research.

¹The HGP, sponsored by the Department of Energy and National Institutes of Health Genome Programs, is the national coordinated effort to characterize all human genetic material by determining the complete sequence of the DNA in the human genome. The HGI is a separate initiative, and is an international research program for the creation of detailed genetic and physical maps for each of the twenty-four different human chromosomes and the elucidation of the complete DNA sequence of the human genome.