INTELLECTUAL PROPERTY

Intellectual Property Landscape of the Human Genome

Kyle Jensen and Fiona Murray*

Gene patents are the subject of considerable debate and yet, like the term "gene" itself, the definition of what constitutes a gene patent is fuzzy (*1*). Nonetheless, gene patents that seem to cause the most controversy are those claiming human protein-encoding nucleotide sequences. This category is the subject of our analysis of the patent landscape of the human genome (*2*).

Enhanced online at www.sciencemag.org/cgi/content/full/310/5746/239

Critics describe the growth in gene sequence patents as an intellectual property (IP) "land grab" over a finite number of human genes (*3*, *4*). They suggest that overly broad patents might block follow-on research (*5*). Alternatively, gene IP rights may become highly fragmented and cause an anticommons effect, imposing high costs on future innovators and underuse of genomic resources (*6*). Both situations, critics argue, would increase the costs of genetic diagnostics, slow the development of new medicines, stifle academic research, and discourage investment in downstream R&D (*7*–*11*).

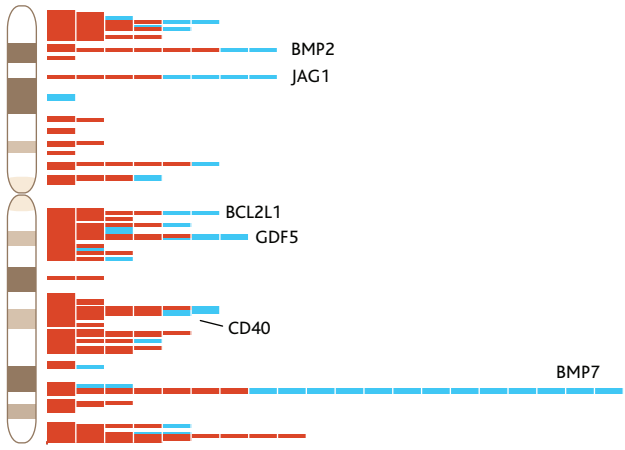
In contrast, the classic argument in support of gene patenting is that strong IP protection provides incentives crucial to downstream investment (*12*, *13*) and the disclosure of inventions. Patents are also regarded as the cornerstone of vibrant markets for ideas (*14*) and central to the biotech boom of the 1980s and 1990s (*15*).

Policy-makers are hampered by the lack of empirical data on the extent of gene patenting. Most analyses have relied on anecdotal evidence (*11*, *16*–*18*) and empirical analyses have been hindered by (i) limited (and poorly defined) coverage of DNA sequence patents (*17*, *19*); (ii) difficulty separating patents that claim gene sequences per se from those merely disclosing DNA sequences (*20*–*22*); and (iii) distinguishing patents on the human genome from those on other species (*23*).

Our detailed map was developed using bioinformatics methods to compare nucleotide sequences claimed in U.S. patents to the human genome. Specifically, this map is based on a BLAST (*24*) homology search linking nucleotide sequences disclosed and claimed in granted U.S. utility patents to the set of protein-encoding messenger RNA transcripts contained in the National Center for Biotechnology Information (NCBI) RefSeq (*25*) and Gene (*26*) databases. This method allows us to map gene-oriented IP rights to specific physical loci on the human genome (*27*) (see figure, right). Our approach is highly specific in its identification of patents that actually claim human nucleotide sequences. However, by limiting the search to patents using the canonical "SEQ ID NO" claim language we do not consider claims on genes defined through amino acid sequences. (See table S1 for a sensitivity analysis.)

Our results reveal that nearly 20% of human genes are explicitly claimed as U.S. IP. This represents 4382 of the 23,688 of genes in the NCBI's gene database at the time of writing (see figure, right). These genes are claimed in 4270 patents within 3050 patent families (*28*). Although this number is low compared with prior reports, a distinction should be made between sequences that are explicitly claimed and those that are merely disclosed, which outnumber claimed sequences roughly 10:1. The 4270 patents are owned by 1156 different assignees (with no adjustments for mergers and acquisition activity, subsidiaries, or spelling variations). Roughly 63% are assigned to private firms (see figure, above). Of the top ten gene patent assignees, nine are U.S.-based, including the University of California, Isis Pharmaceuticals, the former SmithKline Beecham, and Human Genome Sciences. The top patent assignee is Incyte Pharmaceuticals/Incyte Genomics, whose IP rights cover 2000 human genes, mainly for use as probes on DNA microarrays.

Although large expanses of the genome are unpatented, some genes have up to 20 patents asserting rights to various gene uses and manifestations including diagnostic uses, single nucleotide polymorphisms (SNPs), cell lines, and constructs containing the gene. The distribution of gene patents was nonuniform (see figure, page 240, top right): Specific regions of the genome are "hot spots" of heavy patent activity, usually with a one-gene-many-patents scenario (see figure, below). Although less common, there were cases in which a single patent claims many genes, typically as complementary DNA probes used on a microarray (see figure, p. 240, bottom).



Physical mapping of patent activity on chromosome 20, divided into 300-kb segments. Each horizontal bar represents a unique patent claiming a gene sequence located in that region. Orange represents the number of unique patent families in a region (*28*). Labels show the loci of highly patented genes (see table S1).

BMP7, an osteogenic factor, and CDKN2A, a tumor suppressor gene, were the most highly patented genes in the genome [their sequences were each claimed in 20 patents (table S2)]. The patents on CDKN2A are distributed between nine different assignees and, collectively, claim all three splice variants of the gene. Nearly all of these patents are directed toward diagnostic applications. In contrast, the patents on BMP7 are for the use of BMP7 proteins in implants to stimulate bone growth. However, a number are directed towards more speculative utilities, such as drug-screening probes, which suggests a strategy of "science-

K. Jensen is in the Department of Chemical Engineering, F. Murray is in the Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

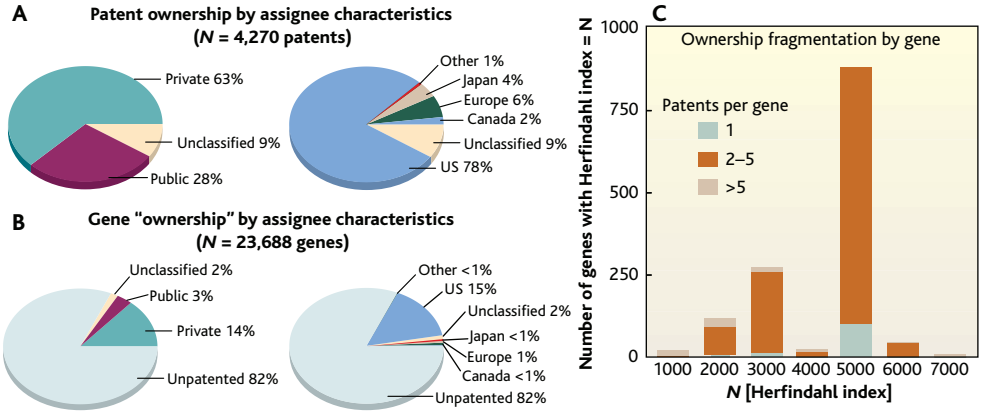
*Author for correspondence. E-mail: fmurray@mit.edu

based" rather than "disease-based" patenting.

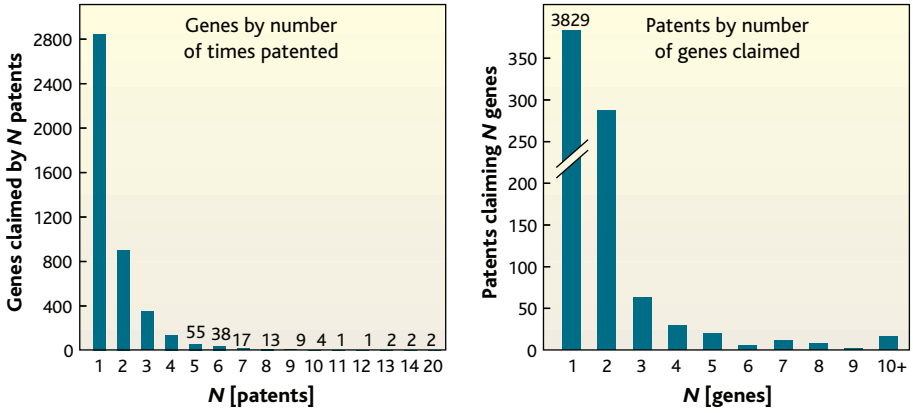
Unsurprisingly, other heavily patented genes tended to have relevance to human health and diseases: e.g., BRCA1 (breast cancer), PIK3R5 (diabetes), and LEPR (obesity). Of the 291 cancer genes reviewed by Futreal *et. al.* (*29*), 131 are patented—significantly more than expected for a random sample of genes ($P = 1.2^{-32}$ based on binomial distribution). Moreover, these genes contain a higher number of patents per gene than expected by chance ($P = 9.4^{-11}$ based on a chi-squared test) (*30*).

Of the 4000+ patented genes, at least 3000 have only a single IP rights holder. For the remainder, we examined whether IP ownership was fragmented by constructing a measure based on the Herfindahl index (*31*) (see figure, top right; part C). The two genes with the most fragmented ownership were PSEN2, the amyloid precursor protein (8 assignees for 9 patents), and BRCA1, the early onset breast cancer gene (12 assignees for 14 patents). Such fragmentation raises the possibility that innovators may incur considerable costs securing access to genes via structuring complex licensing agreements.

Our analysis suggests a number of avenues for further research: It would be valuable to examine whether current practice in patent examination has allowed multiple conflicting patents on the same gene. In addition, genes with multiple patents and IP owners provide a valuable context in which to explore the variety of arrangements used to facilitate or block access to gene-based research and the impact of these arrangements on future innovators. Finally, whereas our study includes only protein-coding genes, future studies should characterize the nature and extent of the rapidly growing IP surrounding non–protein coding components of the human genome, such as microRNAs, ribozymes, and cis-regulatory elements.



Patent and gene ownership characteristics. (**A**) and (**B**) Distribution of gene patent assignees "public" includes governments, schools, universities, research institutions, and hospitals. (A) Ownership breakdown for the 4270 human gene patents. Fractional ownership is based on the number of assignees on a single patent or the number of patents on a gene. (B) "Ownership" breakdown of the genes in the human genome. (**C**) The fragmentation of gene ownership by the Herfindahl index, rounded to the nearest 1000. (*31*). (The 3002 genes with an index of 10,000 are not shown; those for 8000 to 9000 would not be visible on the graph.) The assignee names were used as listed on the patents by the European Patent Office. As such, the Herfindahl indices are likely to overestimate the "true" fragmentation because they do not reflect assignee name changes, mergers, acquisitions, splits, partnerships, or other events that usually lead to a consolidation of IP rights.



Global characteristics of the patent map. (**Left**) Distribution of genes by the number of times they are patented. (**Right**) Distribution of patents by the number of unique genes they claim.

References and Notes

1. E. F. Keller, *The Century of the Gene* (Harvard Univ. Press, Cambridge, MA, 2000).
2. We use the term "gene" to mean a set of cotranscribed protein-encoding exons. Here, "gene patent" means a patent disclosing and claiming a human gene sequence or some fraction thereof.
3. D. Crease, G. Schlich, *Nat. Rev. Drug Discov.* **2**, 407 (2003).
4. T. Hollon, *Nat. Med.* **6**, 362 (2000).
5. S. Scotchmer, *J. Econ. Perspect.* **5**, 29 (1991).
6. M. A. Heller, R. S. Eisenberg, *Science* **280**, 698 (1998).
7. U. S. Patent and Trademark Office, *Fed. Regist.* **66** (4), 1092 (January 2001).
8. Nuffield Council on Bioethics, "The ethics of patenting DNA" (Tech. rep., Nuffield Council on Bioethics, London, UK, 2002).
9. T. Caulfield, E. Gold, M. Cho, *Nat. Rev. Genet.* **1**, 227 (2000).
10. L. Andrews, *Nat. Rev. Genet.* **3**, 803 (2002)
11. S. M. Thomas, M. M. Hopkins, M. Brady, *Nat. Biotechnol.* **20**, 1185 (2002).
12. M. Stott, J. Valentine, *Nat. Rev. Drug Discov.* **3**, 364 (2004).
13. J. J. Doll, *Science* **280**, 689 (1998).
14. J. Gans, S. Stern, *J. Econ. Manage. Strat.* **9**, 485 (2000).
15. R. Levin, A. K. Klevorick, R. R. Nelson, S. G. Winter, *Brookings Pap. Econ. Act.* **3**, 783 (1987).
16. R. Eisenberg, *C. R. Biol.* **326**, 1115 (2003).
17. M. Stott, J. Valentine, *Nat. Biotechnol.* **21**, 729 (2003).
18. J. P. Walsh, A. Arora, W. M. Cohen, *Science* **299**, 1021 (2003).
19. G. Xu, A. Webster, E. Doran, *World Patent Inform.* **24**, 95 (2002).
20. G. Dufresne, M. Duval, *Nat. Biotechnol.* **22**, 231 (2004).
21. Georgetown University Kennedy Institute of Ethics, DNA patent database (http://dnapatents.georgetown.edu/).
22. National Genome Information Center, Patome database (http://bach.ngic.re.kr:8000/patent/en/).
23. G. Dufresne, L. Takacs, H. C. Heus, J.-J. Codani, M. Duval, *Nat. Biotechnol.* **20**, 1269 (2002).
24. S. Altschul, G. Warren, W. Miller, E. Myers, D. Lipman, *J. Mol. Biol.* **215**, 403 (1990).
25. K. D. Pruitt, T. Tatusova, D. R. Maglott, *Nucleic Acids Res.* **33** (database issue), 501 (2005).
26. D. Maglott, J. Ostell, K. D. Pruitt, T. Tatusova, *Nucleic Acids Res.* **33** (database issue), 54 (2005).
27. Materials and methods are available as supporting material on *Science* Online.
28. We use the definition of patent family recommended by the International Patent Documentation Center: Any two patents linked directly or indirectly by a priority document are in the same family.
29. P. A. Futreal *et al.*, *Nat. Rev. Cancer* **4**, 177 (2004).
30. We found similar patenting rates for the 1456 genes listed in the Online Mendelian Inheritance in Man (*32*) with a well-characterized association to disease phenotypes (517 of 1456, P value = 1.6^{-67}).
31. The Herfindahl index is the sum of the squares of the patent shares (in percentage terms) of each patent assignee (range 0 to 10,000), where 10,000 represents a "monopolist" with 100% of patents owned by one assignee and low numbers representing more fragmentation.
32. Online Mendelian Inheritance in Man, OMIM, McKusick-Nathans Institute for Genetic Medicine, Johns Hopkins University (Baltimore, MD) and National Center for Biotechnology Information, National Library of Medicine (Bethesda, MD), 2000.

Supporting Online Material
www.sciencemag.org/cgi/content/full/310/5746/240 DC1

10.1126/science.1120014