"Open and Collaborative" Biomedical Research: Theory and Evidence

Arti K. Rai*

Mounting empirical evidence suggests that biomedical research has, over the last 25 years, become increasingly proprietary[1] and secretive.[2] Given the cumulative nature of research, this trend has raised fears that future progress may be impeded by access and licensing difficulties.[3] One important response has involved calls for improving access by requiring publicly-funded scientists and research institutions to put data and certain types of research tools into the public domain or, at a minimum, to license them widely and nonexclusively at a reasonable fee. [4] This emphasis takes the current organizational structure of research as a given, but seeks to reduce the intensity of exclusionary behavior associated with the research. A response that is perhaps more dramatic has begun to emerge, however. Public funding bodies, prominent scientists, and even some

Deleted: last 25 years,

Formatted

* Professor of Law, Duke Law School. I thank Stuart Benjamin, Jamie Boyle, Iain Cockburn, Wesley Cohen, and Bob Hahn for helpful comments on earlier versions of this paper. Portions of this paper are drawn from an Essay that will be published in Robert Hahn, ed., INTELLECTUAL PROPERTY RIGHTS IN FRONTIER INDUSTRIES: BIOTECHNOLOGY AND SOFTWARE (forthcoming AEI-Brookings Press 2005). The research that is the basis for this paper is funded in part by NIH/DOE Grant # ___.

[1] *See* John P. Walsh, Ashish Arora, and Wesley M. Cohen, *Effects of Research Tool Patents and Licensing on Biomedical Innovation*, in WESLEY M. COHEN AND STEPHEN A. MERRILL, EDS., PATENTS IN THE KNOWLEDGE BASED ECONOMY 285 (National Academies 2003); Arti Rai & Rebecca Eisenberg, *Bayh-Dole Reform and the Progress of Biomedicine*, 66 LAW & CONTEMPORARY PROBLEMS 289 (2003) (discussing increased concentration of proprietary rights in upstream biomedical research).

[2] *See* Eric G. Campbell et al., *Data Withholding in Academic Genetics: Data from a National Survey*, 287 JAMA 473 (2002); Jeremy Gruschow, *Measuring Secrecy: A Cost of the Patent System Revealed*, 33 J.LEGAL STUD. 59 (2004); John Walsh and Wei Hong, *Secrecy Is Increasing in Step With Competition*, 422 NATURE 801 (2003) (empirical findings indicating increased secrecy in biomedical research). I discuss connections between increases in proprietary rights and increases in secrecy *infra* __.

[3] *See, e.g.*, Michael Heller & Rebecca Eisenberg, *Can Patents Deter Innovation? The Anticommons in Biomedical Research*, 280 SCIENCE 698 (1998) (discussing potential problems caused by proliferating rights); Arti Rai & Rebecca Eisenberg, Bayh-Dole Reform and the Progress of Biomedicine, 66 LAW & CONTEMPORARY PROBLEMS 289 (2003) (discussing potential problems caused by broad rights as well as proliferating rights)

[4] For examples of such calls for access, *see, e.g.*, National Research Council, Sharing Publication-Related Data and Materials: Responsibilities of Authorship in the Life Sciences (National Academies Press 2003); Department of Health and Human Services, National Institutes of Health, Principles and Guidelines for Recipients of NIH Research Grants and Contracts on Obtaining and Disseminating Biomedical Research Resources: Final Notice, 64 Fed. Reg. 72,090, 72,093 (Dec. 23, 1999) (research tools)

pharmaceutical firms have taken steps in the direction of what might be called "open and collaborative"[5] science. Open and collaborative projects not only disavow exclusionary behavior but they also move beyond the traditional small-lab/small-firm based structure of biomedical research.

The rise of arguments for open and collaborative biomedical research has coincided with two prominent phenomena: 1) the increased importance of computation in such research[6]; and 2) the well-documented emergence of so-called "open source" methods of innovation in computation-heavy areas of research and development, primarily software. In some recent cases, the modeling on open source software has been quite explicit – for example, the federally funded haplotype mapping project, which aims to create a database that catalogues all human genetic variation, has adopted a licensing policy that is self-consciously modeled on the "copyleft" system of open source software licensing.[7] Under the copyleft version of open source development, users can access source code freely, but such access is conditioned on the user making his or her improvements to such information available under the same conditions.

Although a few commentators have discussed briefly the application of open source-type principles to biomedical research,[8] they have not analyzed carefully how the model is actually being used. In this paper, I draw upon an ongoing, multi-year empirical

---

[5] The term "open and collaborative projects" was recently invoked in a letter to the World Intellectual Property Organization ("WIPO") urging WIPO to hold an exploratory meeting on these types of projects. *See* Letter from Sixty-Eight Scientists and Economists to Kamil Idris, Director General of the World Intellectual Property Organization, July 7, 2003, available at www.cptech.org/ip/wipo/kamil-idris-7July2003.pdf. That letter does not specifically define its use of the term. This paper's definition is set out in the text.

[6] *See, e.g.* William Jorgensen, *The Many Roles of Computation in Drug Discovery*, 303 SCIENCE 1818 (2004).

[7] See International HapMap Project Public Access License, available at www.hapmap.org/cgi-perl/registration (acknowledging model of GNU General Public License)

[8] *See, e.g.*, Dan Burk For a preliminary empirical investigation in the journalistic literature *see* Kenn Cukier

inquiry into the role of intellectual property in computational biology and associated efforts[9] to evaluate the extent to which the open and collaborative research model may promote socially desirable biomedical innovation – that is, innovation that produces marginal health benefits in excess of its marginal costs.[10]

Whether the open and collaborative model is likely to promote such innovation, either as an absolute matter or relative to more traditional models, is a difficult question to answer. Indeed, because the model is quite fresh, and the time delay before research on this model can be translated into end products is long, empirical demonstration of the model's virtues and vices is, at this stage, probably impossible. Nonetheless, the growing use of the model in three prominent contexts: bioinformatics software; databases; and "wet lab" systems biology research affords the opportunity for some tentative commentary.[11]

---

[9] Arti K. Rai and Bhaven Sampat, *Intellectual Property and Computational Biology: An Empirical Inquiry* (series of working papers).

[10] More specifically, I mean innovation that achieves a marginal benefit of one additional quality-adjusted life year (QALY) at a marginal cost of less than $50,000 to $100,000. Although marginal cost per QALY gained is a standard metric for evaluating technology in the health and environmental arenas, innovation scholars typically do not engage in direct evaluations of technology. Rather, they tend to assume that the combination of market demand and well-calibrated property rights will produce the appropriate rate and direction of innovation. My use of the QALY metric thus merits some explanation. In health policy analysis, QALYs have achieved currency because market purchasing decisions are not necessarily a satisfactory measure of social welfare. Most health care is paid for through insurance, and many studies show that insured individuals have incentives to consume more health care than is cost-beneficial. This phenomenon of insurance-induced moral hazard appears to have led to some technological innovation that is not cost-effective. For example, it appears that many profitable pharmaceuticals represent only marginal improvements over existing alternatives. *See, e.g.*, Arti K. Rai, *The Information Revolution Reaches Pharmaceuticals: Balancing Innovation Incentive, Cost and and Access in the Post-Genomics Era*, 2001 U.ILL.L. REV. 173 (discussing demand-side reforms that might channel innovation in cost-effective directions). Measurements based on QALYs also do not discriminate on the basis of ability to way. Thus, for those (myself included) who believe that some level of health care redistribution is morally justified, they are a more appropriate measure of social welfare than willingness to pay.

[11] This paper addresses only collaborative projects in which participants are contributing *intellectual* efforts (either alone or, more typically, in combination with physical resources). It does not address projects in which volunteers are donating only "excess" physical resources (e.g. CPU cycles). For a thorough discussion of such projects in the area of distributed computing, *see* Yochai Benkler, *Sharing Nicely*.

In software and database production, where the innovative task can be subdivided into units that are amenable to being pursued independently (and then aggregated, generally via the Internet) the model is intuitively applicable. Additionally, the available empirical evidence indicates that open and collaborative approaches represent a low transaction cost mechanism for producing biological software and data that is not only useable but potentially superior to its proprietary counterparts. The resulting public domain status for this software and data will also reduce access and licensing impediments to follow-on wet lab research. On the other hand, at least in the context of databases, the open and collaborative approach will require public funding and will need to be restricted, at least in the initial information production stage, to a limited number of players. Publicly funded databases will also, in all likelihood, undermine the viability of private database firms. Perhaps most importantly, information production in areas where the innovative task is relatively well-understood/codified and thus can be partitioned into smaller pieces – that is, made modular[12] – could be accomplished not only through open source approaches but also through markets. Though market-based production is likely to have transaction costs that are somewhat higher than open source production, codification and subsequent modularization tends to reduce such costs to levels that allows many market-based projects to go forward.[13]

---

[12] "Modularity" is a concept that is central to software development. Most software development proceeds according to this model. As we will see, it can also be readily applied in other digital contexts, for example genomic databases. In the context of databases, it is important to emphasize that modularity applies not only to dividing the data production task (i.e. different participate sequence different chromosomes) but also to the standardized research inputs – such as automated laser sequencing machines – that are used to conduct the task.

[13] *See, e.g.*, ASHISH ARORA ET AL., MARKETS FOR TECHNOLOGY: THE ECONOMICS OF INNOVATION AND CORPORATE STRATEGY 6-7 (2001) (noting that although R&D has historically been integrated in large firms, the codification of previously tacit knowledge has reduced the transaction costs of exchange and made possible the rise of small firms that market technology); *see also* Robert Merges, *Intellectual Property and the Costs of Commercial Exchange*, 93 MICH.L.REV. 1570, 1573-74 (1995) (noting that

In contrast, there is reason to believe that significant levels of innovation in inchoate areas – that is, areas where the relevant information has not been well-characterized – are very difficult to accomplish outside large firms. Given the potential for large hierarchies to discourage creative research, confining innovation in these areas to large firms may be undesirable. Moreover, even a single large firm may be unable to hedge the risk associated with conducting substantial research in inchoate areas. And multi-firm collaborations are likely to get bogged down in disputes over ownership of intellectual property rights. Thus the least intuitive, but potentially most exciting, of the a publicly-subsidized open and collaborative approach may involve the inchoate area of wet lab systems biology: in this context, the model may allow a more coordinated and comprehensive attack on complex, previously intractable problems than does traditional small lab biology. The dearth of knowledge regarding systems biology is an important reason many promising drug candidates, particularly for complex diseases influenced by multiple genes, currently fail in preclinical or clinical trials.[14] Hence open and collaborative approaches have the potential to address the innovation drought currently facing the biopharmaceutical industry.[15]

---

intellectual property rights facilitate the division of innovative labor). Merges appears to assume that intellectual property rights will attach to information with relatively clear boundaries. Indeed, various doctrines of patent law aim to foster such clarity. Of course, most information can not be made perfectly divisible and excludable. For this reason, Kenneth Arrow and others have long argued that patents provide inadequate incentives to engage in R&D. *See* discussion *infra* __.

[14] Most obviously, if a drug targets only one gene or protein, then it will not be effective against a multi-gene disease. But even drugs for single-gene diseases often fail because they end up affecting other genes and proteins and having unexpected side effects. *See* Caitlin Smith, *A Question of Biology*, 428 NATURE 225, 231 (2004) (noting comment by scientist that "[I]n the past, a simplistic view was, by necessity, taken, which resulted in many drugs failing in preclinical or clinical trials for lack of efficacy or side effects. The new approach must account for this complexity. The holistic systems biology approach to research will be necessary to overcome this challenge.") See also discussion *infra* __.

[15] Iain M. Cockburn, *The Changing Structure of the Pharmaceutical Industry*, 23 HEALTH AFFAIRS 10, 11 (2004); Robert F. Service, *Surviving the Blockbuster Syndrome*, 303 SCIENCE 1796 (2004) (discussing low number of new chemical entities approved in 2002). Of course, the moral hazard problems created by health insurance, *see supra* n.11, also make it rational for pharmaceutical firms to focus their

Certain applications of the open and collaborative model do, however, raise concerns. One important concern involves the possibility of reduced incentives for development and commercialization as research moves downstream, towards the chemical compound that will be the drug candidate. As various empirical studies have documented, patents on chemical compounds are critical for purposes of recouping the large costs associated with pre-clinical and clinical R&D.[16] To preserve development and commercialization opportunities, there is reason to be quite cautious about deploying copyleft licensing, at least outside the context of software. Several institutional concerns also bear mention. First, given conventional biology's focus on peer-reviewed print publications, the Web-based publication model used by some open and collaborative projects may make it difficult for such projects to attract attention and promising young investigators. Second, university technology transfer offices (TTOs), which are relevant given the academic locus of most open and collaborative projects, may represent an obstacle.

Part I of this paper gives economic and institutional background on biopharmaceutical innovation, with an eye towards highlighting the innovation challenges to which the open and collaborative model could respond. Part II discusses how open collaboration has operated in the context of software and of some other Web-

---

energies on marginal patentable improvements rather than risky breakthrough drugs. So improving the research model is a necessary but not sufficient condition for the production of breakthrough drugs.

Another area that could be fruitful ground for open and collaborative research of both the computational and wet lab variety is the collection of contexts where patents clearly do little or no work, either because the relevant compound is already in the public domain or because there is no substantial paying market for a patented product. In these arenas, one major drawback of open research – that it could defeat patents necessary for development and commercialization -- is not a problem. For one such proposed project, *see* Stephen Maurer, Arti Rai, and Andrej Sali, *Finding Cures for Tropical Diseases: Is Open Source An Answer*? __ PUBLIC LIBRARY OF SCIENCE: MEDICINE (forthcoming December 2004). *See also An Open-Source Shot In the Arm*?, THE ECONOMIST, June 10, 2004 (discussing proposal). Because this tropical disease project is not yet operational, however, this paper does not focus on it.

[16] *See, e.g.*, Wesley Cohen et al., *Protecting Their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (or Not*), NBER Working Paper No. 7552 (2000).

based information projects. It also discusses results from empirical investigation of some prominent open and collaborative biomedical research projects. Part III uses these results as well as the theoretical literature to elucidate the extent to which the open and collaborative model may produce socially desirable biomedical innovation. In this Part, I also make recommendations for removing institutional obstacles in those cases where the model may be superior to alternative arrangements.

Part I: The Open and Collaborative Model in Context

A. Innovation in Biopharmaceuticals

For much of the twentieth century, innovation was typically conducted in large, vertically integrated firms. Indeed, many economists argued that private-sector innovation necessarily occurred most smoothly through this model. On this view, the tacit, uncodified nature of much technological knowledge made procuring innovation through markets an activity that was bound to be rife with transaction costs.[17] Additionally, even where knowledge was codified, it was not always, or even generally, a possible subject for patent protection. As a consequence, Arrow's information paradox created significant openings for opportunistic behavior in negotiation.[18] Although trade secrecy protection could of course protect to some extent against opportunistic behavior, trade secrecy provided imperfect protection.

Innovation in biopharmaceuticals was no exception to the large-firm rule. Through a combination of size and monopoly-conferring end product patents, large

---

[17] *See, e.g.*, David J. Teece, *Technological Change and the Nature of the Firm*, in G.Dosi et al., ed., TECHNOLOGICAL CHANGE AND ECONOMIC THEORY (1988); RICHARD NELSON AND SID WINTER, AN EVOLUTIONARY THEORY OF ECONOMIC CHANGE (1982). These discussions obviously apply to innovation the Coasean insight that firms form when the transaction costs of operating in the marketplace are high. See Ronald Coase, *The Nature of the Firm*.

[18] Kenneth Arrow, *Economic Welfare and the Allocation of Resources for Invention* in THE RATE AND DIRECTION OF INVENTIVE ACTIVITY 609, 615 (NBER 1962).

pharmaceutical firms hedged the risk associated with their unsystematic, trial and error-based innovation. To the limited extent that scientific knowledge in the biomedical arena was codified, such codification did not generally confer patent protection. The Supreme Court interpreted the utility doctrine of patent law as a bar against patenting information that did not have a specific application for end users (as contrasted with researchers).[19] Moreover, government patent policy made it difficult for the academic institutions that were conducting the relevant research to seek patent protection.[20]

Until the 1970s, then, academic biomedical science generally steered clear of both firms and markets. This is not to say that it adhered fully to the norms of scientific communalism famously described by sociologists like Robert Merton.[21] The authors of a recent study that reanalyzes data from the 1960s argue that, even in 1966, experimental biologists were more reluctant than scientists in other fields to discuss ideas freely outside their individual lab.[22] Indeed, this tradition of greater secrecy may have made the biological sciences more hospitable territory for subsequent patenting activity than (for example) high-energy physics or computer science. Nonetheless, secrecy was fueled by

---

[19] *Brenner v. Manson*

[20] *See, e.g.,* Arti K. Rai, *Regulating Scientific Research: Intellectual Property Rights and the Norms of Science*, 94 NW.L.REV. 77, 95 (1999).

[21] Robert K. Merton, *The Normative Structure of Science, reprinted in* ROBERT K. MERTON, THE SOCIOLOGY OF SCIENCE (1973) (arguing that, in general, academic research science is a communal enterprise). Mertonian norms are often characterized as a common property/governance regime of the sort analyzed by Elinor Ostrom and Robert Ellickson. *See* Rai, *Regulating Scientific Research supra* __. These norms also have much in common, however, with a true open access regime. *See* Carol Rose, *The Comedy of the Commons* (discussing such regimes). Because scientific research is generally directed at (and comprehensible to) other scientists only, the open access framework is not immediately apparent. But this framework has become particularly salient in recent months, with Congress now considering a proposal, supported by many citizen advocacy groups, that publicly funded medical research be made freely available to all interested parties six months after publication. Additionally, as discussed further below, *see infra* __, the Mertonian framework certainly requires less intensive governance than copyleft-style licensing.

[22] John P. Walsh and Wei Hong, *Secrecy is Increasing in Step with Competition*, 422 NATURE 801 (2003)

academic competition, not commercial competition. Additionally, levels of secrecy were significantly lower than those that exist today.[23]

By the mid to late 1970s, molecular biology had made significant advances in codification. Recombinant DNA technology and monoclonal antibody research represented two of the first, and most important, such advances. Just as the field was becoming codified, Congress passed legislation that made it easier for both universities and private industry to seek intellectual property rights in such codification. In 1980, Congress passed the Bayh-Dole Act, which specifically encouraged universities to secure patent rights in their federally funded discoveries. In 1982, Congress created the Court of Appeals for the Federal Circuit, which has enhanced patent availability in general, and particularly in the area of biotechnology, by relaxing the utility requirement.[24] Since the passage of Bayh-Dole, universities have focused their patenting activity on biotechnology research.[25] Through exclusive licensing from universities and through their own patenting of research inputs, small firms and startups have also secured strong proprietary positions in biotechnology research.

To the extent one views intellectual property rights as similar to rights in tangible property,[26] these legislative changes could be seen as an example of Harold Demsetz's thesis that property rights that promote optimal development and commercialization will

[23] Walsh and Hong, *supra* __.

[24] *See, e.g.,* In re Brana, 53 F.3d 1560 (Fed.Cir. 1995) (suggesting that utility in research is sufficient for patentability)

[25] *See* Rai and Sampat, *supra* __ (data indicating that in 2000 about 50% of patents held by research universities are in biomedical arena). In 2000, the university presence accounted for about 15% of all biomedical patents. *Id.*

[26] Congressional passage of the Bayh-Dole Act was very much a reflection of this view. *See, e.g.*, H.R. Rep. No. 96-1307, pt. 1, at 3 (1980) (arguing that university patenting and exclusive licensing were necessary to develop inventions to the point of commercial application). *See also* Edmund Kitch, *The Nature and Function of the Patent System*, 20 J.L.& ECON. 265 (1977) (applying property rights theory to patents and discussing how an early grant of rights can guard against both "too much" and "too little"investment in a technological prospect). In contrast, traditional theorists have tended to view intellectual property as an incentive for creating the public good of information.

Formatted

emerge when the opportunity cost of not having such rights exceed the cost of delineation and enforcement.[27] Alternatively, these changes could be seen as a lesson in interest group theory – universities and small firms saw possibilities for revenue enhancement and pressured Congress into passing the legislation necessary to lower barriers to patenting.[28]

In either event, the pace of codification and patenting has only intensified over the last ten years. With the infusion of genomic and proteomic information, pharmaceutical firms all aim to produce drugs by systematically testing their drug compound libraries on genomic and proteomic "targets." The race to patent such upstream information is intense, both among universities and among small firms and startups. In the case of universities, licensing upstream research produces revenue. For small firms and startups, upstream patents – or exclusive licenses to upstream university patents – appear to be important not only for securing licensing revenues but also for attracting venture capital. Even for research that is not patented, upstream players may attempt to leverage their control over data or tools that can not readily be reproduced to exact reach-through royalties. Although physical property rights over materials, and trade secrecy type protection over data, are less useful than patents,[29] they do provide some protection. For their part, large pharmaceutical firms – once vertically integrated engines of innovation – must now negotiate a complex array of university and small firm proprietary claims on

---

[27] Harold Demsetz, *Toward a Theory of Property Rights*, 57 AM.ECON.REV. 347 (1967). The property rights approach to patents has particularly explanatory force with respect to university patenting. In that case, federal funding obviates the need for patents to perform their traditional role of securing the public good of initial information production. Thus the only justification for patenting publicly funded research is property theory, which focuses on development and commercialization.

[28] *See* Saul Levmore, *Two Stories About the Evolution of Property Rights*, JOURNAL OF LEGAL STUDIES S421 (2002) (making a cogent argument that transitions in property regimes can be framed in public interest or public choice terms).

[29] *See supra* __.

research inputs. While some of these claims may be narrow in scope,[30] other claims may be broader.[31] Significantly, with the increasing importance of computation, particularly software, in biomedical research, software is now another category of patented research tool that may add to upstream complexity.

B. Vertical "Dis-Integration" and Calls for Access

As noted, property rights on codified research inputs have fostered the creation of small firms that market such inputs. To the extent that small firms may be more innovative than large firms[32] -- and thus produce research inputs better and faster than large firms – this change could be positive.[33] Additionally, to the extent that research inputs are licensed widely to interested downstream developers, the creation of a market for such inputs could conceivably reduce duplicative research and increase downstream competition.[34] On the other hand, as Coase might predict, the move away from the vertically integrated firm has increased transaction costs substantially. Although such increases do not appear to have caused ongoing projects to stop,[35] there is some evidence that broad patents on research inputs do limit follow-on research.[36] There is also evidence of research delay and of firms avoiding research areas where there are

---

30 *See, e.g.*, University of Rochester v. G.D. Searle, 358 F.2d 916 (2004); Regents of the Univ. of Cal. v. Eli Lilly, 119 F.3d 1559 (Fed. Cir. 1997) (striking down broad patents on biomedical research)

31 Amgen v. Hoechst Marion Roussel, Inc., 314 F.3d 1313 (Fed. Cir. 2003) (upholding broad biomedical patent).

32 Empirical evidence regarding the superior innovative capacities of small firms is mixed. *See* Zoltan J.Acs and David B. Audretsch, *Innovation in Large and Small Firms: An Empirical Analysis*, 78 AMERICAN ECONOMIC REVIEW 678 (1988) (finding that in 21 of 35 industries, large firms were more innovative than small firms). In the pharmaceutical industry, however, biotechnology firms have historically been the engines of innovation. Some firms, perhaps most notably GlaxoSmithKline, have moved to emulate the small firm model by setting up in-house labs that have substantial autonomy.

33 For an argument along these lines, *see* Arora et al. at 7.

34 *Id.*

35 Walsh, Arora, Cohen, *supra* __.

36 *Id.* at __.

significant patent positions.[37] Additionally, in at least two recent cases, academic institutions with research input patents (or their small firm licensees) have tried to "hold-up" pharmaceutical companies that are producing drugs that allegedly infringe the patent on the research input.[38]

Problems associated with even purely academic access appear to have become more prevalent.[39] With respect to data and materials to which researchers need physical access,[40] both increased commercialization and increased scientific competition between labs have contributed to access difficulties. In a survey conducted in the mid-1990s, 20% of academic respondents in the life sciences reported having delayed publication for more than six months, either for reasons related to commercialization (for example, the need for secrecy before a patent application is filed or the desire for trade secrecy)[41] or because of scientific competition.[42] In a survey of genetics researchers conducted in 2000, 47% of respondents reported having had a request for data and materials related to published research denied. This 47% figure represented a substantial increase over prior surveys. In 21% of cases, such denials caused the academic investigator requesting access to

---

37 *Id; see also* Josh Lerner, *Patenting in the Shadow of Competitors.*

38 University of Rochester v. Searle; Ariad v. Eli Lilly. In both of these cases, Bayh-Dole property rights theory had limited application. The pharmaceutical company in question was happy to engage in development and commercialization without the need for an exclusive license to the relevant upstream patent.

39 Eric G. Campbell et al., *Data Withholding in Academic Genetics: Data From a National Survey*, 287 JAMA 473, 478 (2002). Thirty-five percent of geneticists said that sharing had decreased during the 1990s, while only 14% said that sharing had increased.

40 For materials that academic researchers can readily reproduce on their own, many researchers have been able to secure an informal regime of price discrimination by simply ignoring patent rights. Walsh, Arora, & Cohen at __. Whether this informal price discrimination regime will survive the decision, enunciated by the Court of Appeals for the Federal Circuit in *Madey v. Duke*, to eliminate any possibility of a patent infringement exemption for research is unclear.

41 *See* 35 U.S.C. 102(b) (establishing that an invention can not be patented if it has been disclosed publicly more than one year before a patent application is filed). Retaining the ability to file in foreign jurisdictions may require even more strict secrecy.

42 1997 Blumenthal study, JAMA

abandon a promising line of research.[43] Once again, survey respondents who admitted denying access cited both commercial considerations and academic competition.[44]

Finally, and perhaps most importantly, even as codification of upstream knowledge, and upstream patenting, by academic labs and small firms has increased, this codification has not led to significantly greater understanding of larger biological systems. In other words, although we have identified thousands of individual genes, and may even know something about the work done by the proteins produced by those genes, we have only a glancing knowledge of how these proteins interact with each other (as well as external stimuli such as promising drugs) in the complex signaling and regulatory pathways that determine how a particular cell in a given organism functions. For this reason, the biopharmaceutical industry is facing something of an innovation drought, particularly with respect to drugs for diseases whose etiology is influenced by multiple genes and their associated proteins.

As the web of upstream proprietary rights and secrecy has grown, various public and private sector groups have made calls for greater access to research tools. In 1999, for example, the National Institutes of Health ("NIH") issued a set of guidelines urging federally funded institutions to refrain from patenting and exclusive licensing in cases

---

[43] Campbell, *supra* note __, at __.

[44] *Id.*; *see also* Jeremy Gruschow 2004 (finding more secrecy even when patents were not sought). The precise reason for increased academic competition is not clear. Such competition may stem from the reality that larger numbers of post-graduate doctoral students are competing for a stable number of tenure-track jobs. Alternatively (or additionally), increases in commercial activity may have led to a general breakdown in traditions of reciprocity and survey respondents may be attributing their non-cooperative behavior to the (relatively) inoffensive motive of academic competition. For example, a scientist may refuse to provide research materials, even when she is not seeking commercial protection herself, because she believes her colleagues will not reciprocate in the future due to commercial obligations. *See* Dan Kahan, *The Logic of Reciprocity: Trust, Collective Action, and Law*, 102 MICH.L.REV. 71, 90-98 (2003) (discussing the fragility of a cooperative equilibrium both generally and with specific reference to academia). She may then categorize this failure to share as motivated by academic competition. For a general argument that monetary incentives can "crowd out" non-monetary incentives, *see* Yochai Benkler, *Coase's Penguin*, 112YALE L.J. 369, 404-05 (2002).

where the biomedical research tool in question is "a broad, enabling invention that will be useful to many scientists, or multiple companies in developing multiple products . . ."[45] The guidelines also urge universities to exchange all unpatentable research tools freely, both within the academic community and with industry. A recent report by the National Research Council focuses specifically on the problem of secrecy, arguing that professionally norms oblige scientists to share the data and research materials necessary for replicating published work.[46] The pharmaceutical company Merck has even attempted to preempt patent claims by putting partial expressed gene sequences into the public domain.

II. Beyond Access: Open and Collaborative Research

Given the problems that may be created by new proprietary rights, and are almost certainly created by ever-increasing levels of secrecy, calls for access are important. Indeed, in many cases, access to research inputs will be essential if the open and collaborative model is to succeed. Nonetheless, the model moves beyond Mertonian access arguments in that it explicitly requires scientists not only to be open but also to work closely with others outside their own small lab or small firm. In this section, I describe some prominent open and collaborative biomedical research projects. Because many proponents of this research invoke the example of open source software, and because some of this research is in fact itself open source software, I first discuss the open source model.

A. The Open Source Model

---

45 *See supra* note 5.

46 *See supra* note 5.

Open source software development has close ties to the norm-based Mertonian framework for conducting scientific research. Indeed, the open source movement originated in a communal "hacker" culture that prevailed in certain academic and federal laboratories in the 1960s and 1970s. At that time, packaged software was rare and individuals freely exchanged software and underlying source code for purposes of modification and improvement. Such exchange was facilitated with the creation of the ARPANET network, which eventually expanded into the Internet. The transaction cost-lowering properties of the Internet allowed Mertonian norms to operate more effectively and in larger, more disparate groups than they ordinarily operate.[47]

Currently, open source software production is a large and heterogeneous phenomenon encompassing approximately 26,000 projects.[48] All open source projects require that the licensee receive, and be able to redistribute,[49] source code. Beyond this basic requirement, open source licenses (of which there are more than 30 varieties) can roughly be divided into two different categories: the category of copyleft or "GPL" licenses that require licensees who make improvements to the software to make those improvements publicly available on the same terms that they received the software;[50] and a second category that discloses source code but essentially imposes few if any requirements on recipients. An argument often made in favor of copyleft licenses is that,

---

47 Cf. ROBERT ELLICKSON, ORDER WITHOUT LAW: HOW NEIGHBORS SETTLE DISPUTES 3 (1990); ELINOR OSTROM, GOVERNING THE COMMONS 88-89 (1990) (noting that enduring norms operate where the group is relatively small and cohesive).

48 See www.sourceforge.net (registry of open source projects)

49 Bruce Perens, *The Open Source Definition*, available at http://perens.com/articles/osd.html.

50 Although Richard Stallman and some others argue that copylefted software should be called "free" software, this paper uses the term "open source" to encompass copylefted software.

by preventing private appropriation of volunteer labor, such licenses provide an incentive for volunteers to contribute in the first instance.[51]

Notably, although most open source projects appear to use GPL-style licenses,[52] some recent empirical studies indicate that such licenses are significantly more common in projects that are consumer-oriented. In contrast, projects that are geared towards technical users (for example, scientific and engineering programs) are less likely to have restrictive licenses.[53] Moreover, GPL-style licenses appear to be common in projects with many contributors, each of whom contribute a small amount of code to the project. In contrast, in projects with non-restrictive licenses, the number of contributors is smaller but output per contributor (as measured in source code lines) is significantly higher .[54] Taken together, these empirical results indicate the open source software production comes in two categories that divide roughly according to license type. One category of open source production is conducted by, and directed to, professional scientists and engineers. In this category, the number of contributors is relatively small and legal restrictions may little if any role. This category appears to be closest to the Mertonian model. In the second category, the number of contributors is larger and more heterogeneous, they contribute less code, and law is used in conjunction with norms to prevent defection.[55] Extending Demsetz, one could argue that the latter set of projects is

---

[51] Jonathan Zittrain, *Evaluating Free and Proprietary Software*, 71 U.CHI.L.REV. 265, 279 (2004). *Cf.* Benkler, *supra* note __ (discussing "crowding-out" motivations).

[52] According to Josh Lerner and Jean Tirole, approximately 70% of the 25,729 projects found at www.sourceforge.net used GPL-style licenses. Lerner and Tirole, *The Scope of Open Source Licensing*, 2002.

[53] Lerner and Tirole (2002)

[54] Fershtman and Gandal (2004). Note, however, that both the Lerner and Tirole and Fershtman and Gandal studies rely upon SourceForge, which does not host such well-known open source projects as LINUX, APACHE, Perl, and SendMail.

[55] It is important to note, however, that no case involving a restrictive open source license appears to have been litigated to judgment. Open Source Discussion at American Bar Association, Joint Session of

more commercially valuable, and it is therefore efficient to institute more complex governance mechanisms.[56] Alternatively (or additionally), one could argue that Mertonian norms are likely to operate with greater force in small communities of scientists and engineers than in more heterogeneous communities of developers. Hence even a passing reference to law is not necessary. As we will see, the open source software development that is currently taking place within the biomedical research community generally adheres more closely to the Mertonian model than to the copyleft model.[57]

An important divergence from Mertonian norms that is found even within communities that develop scientific and technical software lies in mechanisms for information integration. While the Mertonian model does not posit a specific mechanism for information integration, open source software production, particularly production in large-scale projects, often has a central developer or group of developers who are responsible for evaluating and integrating developments on an ongoing basis. New and modified code that is deemed to be of sufficient quality by the developer may then be

---

Intellectual Property Section and Science and Technology Section, 4/1/04 . There are two recorded instances of litigation brought by holders of copyleft licenses claiming improper propertization of the code. Zittrain, *supra* note __, at 285. Rather, according to one study, the primary vehicle for enforcement is identification and critiquing of violations on on-line mailing lists and bulletin boards. Siobhan O'Mahony, *Guarding the Commons: How Community Managed Projects Protect Their Work* 32 RESEARCH POLICY 1179, 1189 (2003).

[56] *Cf.* Henry Smith, *Exclusion versus Governance: Two Strategies for Delineating Property Rights*, 31 J.Legal Stud. S453 (2002) (arguing that the Demsetzian vision of property should embrace collective governance mechanisms).

[57] In an important article, Yochai Benkler reviews Web-based collaborative projects, including but not limited to software, and makes the general claim that volunteer collaborations are likely to be superior to firms and markets in allocating human creativity where the task is modular; the cost to volunteers of contribution to a single module is low; and where such contributions can be readily filtered and aggregated. Benkler, *Coase's Penguin*, *supra* __. The empirical evidence to date, which reveals the largest outpouring of collaborative activity in software, does suggest modularity, filtration, and aggregation are important. But the requirement that the cost of contribution be low would appear to be most applicable to projects where the product is non-specialized in nature. As discussed in the text, in specialized scientific and technical projects, volunteer contributions tend to be quite substantial. Of course, the number of participants in these more specialized projects is also likely to be smaller than in non-specialized projects. So these projects might not fall into the category of large-scale collaboration on which Benkler focuses.

added to the official version of the code.[58] To some extent, the control exercised by the developer resembles that exercised by firm management. On the other hand, entry and exit from developer status is more fluid than entry and exit from firm management. Thus, particularly for projects that involve significant amount of source code, open source software production could be seen as lying somewhere between Mertonian norms and the firm.[59]

A final respect in which most open source software production, even in scientific and technical areas, appears to differ from Mertonian science (and, as we will see, from open source applications in the biomedical arena) is that it is generally *not* funded publicly.[60] Not only have prominent firms have been built on providing services for open source software[61] but according to one recent study of 287 open source projects, 38% percent of open source software developers make their contributions at work, with the knowledge of their supervisors. Presumably the firms for which these developers work value the specific improvements that the developers make.[62]

Proponents of open source software argue that such software development works in the sense that it produces useable output at a lower cost than conventional proprietary development. Some make the more ambitious claim that this output, which significant

---

[58] *See generally* Eric von Hippel and Georg von Krogh, *Editorial, Special Issue of Open Source Software Development*, 32 RESEARCH POLICY 1149 (2003). The central developers' control of the project is sufficiently high that "forking" of the source code is rare. Eric Raymond, *The Magic Cauldron*, Sections 3-5.

[59] *Compare* David McGowan, *Legal Implications of Open-Source Software*, 2001 U.ILL.L.REV.241 (discussing respects in which open source software production is, and is not, like firm-based production)

[60] According to one survey, only about 7% of open source software developers work in the academic sector. Karim R. Lakhani and Robert G. Wolf, *Why Hackers Do What they Do: Understanding Motivation Effort in Free/Open Source Software Projects*

[61] Perhaps the most prominent example is Red Hat, which provides Linux-related services.

[62] Lakhani and Wolf. In general, contributors to open source projects have a wide variety of intrinsic and extrinsic motivations for contributing – personal enjoyment, sense of community obligation, pay, solving a specific problem, honing skills, and enhancing career prospects. Lerner & Tirole 2001; Benkler (2002) (dividing incentives into monetary, hedonic, and social/psychological)

numbers of independent programmers continually examine for defects, and for the possibility of adding additional features, is likely to be technically superior to closed source output. A small number of technical studies have tested the latter claim. One academic study that compared Linux, Apache, and GCC with their closed-source counterparts appears to buttress claims that open source software may be technically superior. The study determined that open source software had a higher rate of function modification (i.e. fixing of defects) and also added more functions over time.[63] Similarly, Reasoning, Inc., a software inspection service, determined in a 2003 report that the Linux TCP/IP stack had fewer defects than commercially developed TCP/IP stacks. Reasoning, Inc.'s analysis did find, however, that Apache had as many defects as its commercial counterpart. According to the authors of the latter study, this result may be a consequence of the Apache product still being relatively early (as compared with Linux) in the software life cycle. To be sure, given the heterogeneity of both open and closed source software, attempts to generalize from a small number of case studies are perilous. Nonetheless, these studies do show that open source software is a reasonable alternative to closed source. Additionally, because of its reliance on volunteer labor, open source software may in many cases be cheaper than its closed source counterparts.[64]

---

[63] James W. Paulson et al., *An Empirical Study of Open-Source and Closed-Source Software Products*, 30 IEEE TRANSACTIONS ON SOFTWARE ENGINEERING 246 (2004). Interestingly, this study did *not* confirm two other common beliefs about open source software projects – that they succeed because of their simplicity or that they are more modular than closed source projects. *See also* J. Kuan, *Open Source Software as Lead User's Make or Buy Decision* (arguing that the Apache web server, the Linux operating system, and the Gnome user interface had faster rates of bug report resolution than three similar closed source programs).

[64] For firms that use software platforms, however, the cost of the software may be only a small part of the total cost of ownership. In particular, staffing costs for such platforms can be quite high. *See* Alan MacCormack, *Total Cost of Ownership for Software Platforms: Comparing Apples, Oranges, and Cucumbers*. On the other hand, staffing costs are likely to be less important where the software is question is itself directed at a technical audience.

Against this empirical and theoretical background, the remainder of the Article discusses and evaluates open and collaborative projects in the biomedical arena. More specifically, it evaluates the extent to which they are likely to: 1) produce technically competent output; 2) alleviate transaction cost and secrecy problems that may, as discussed earlier, dissuade firms and academics from pursuing promising lines of research; and 3) address scientific challenges, thus far unaddressed, in inchoate but vitally important areas such as system biology.

B. Open and Collaborative Biomedical Research

In this section, I describe various efforts at open and collaborative biomedical research. Because the relevant technical, organizational, and economic considerations are distinct, I treat software, databases, and "wet lab" systems biology as separate categories. Part III then turns to an evaluation of the projects.

1. Bioinformatics Software

Because software design is a skill that relatively few traditional biological researchers have, many bioinformaticians currently come from the computer science community. They bring with them familiarity with open source models. As a consequence, many bioinformatics software projects, particularly small projects where the software is secondary to the biological research, operate under an open source model.[65] Open source is seen as a good mechanism for information dissemination, reduction of duplicative effort, and rapid development of software.[66] By the same

[65] Interview with computational biologist and Bioperl developer Steven Brenner, UC Berkeley.

[66] According to the founding developers of Bioperl, "[a] primary motivation behind writing the toolkit is the authors' desire to focus energies on a solution whose components can be shared rather than duplicating effort . . . In this spirit, we chose to make our code freely available under an open-source license (Open Source Initiative 2001), so that others could extend routines already in the Bioperl library and contribute their own routines as well . . . the open nature of the Bioperl project reduced the time for

token, devotees of open source do not necessarily believe that all bioinformatics software should be open source.[67] Moreover, consistent with the fact that their user audience is highly specialized, most open source bioinformatics projects use non-restrictive licensing.[68]

One important difference between most open source software and open source bioinformatics software is that development of the latter is generally conducted in the academic sector. Moreover, because almost all major research universities require that employee rights in software developed using university resources be assigned to the university,[69] the policy of universities towards open source software development becomes quite relevant.

Preliminary results from interviews with technology transfer offices ("TTOs") at 20 universities that have large software patent, biomedical patent, and/or biomedical research portfolios indicate that most university TTOs have not, at least thus far, been seeking many software patents.[70] The reason is economic: because software licensing typically yields little in the way of licensing revenues, software does not generally merit the cost of a patent filing. To the extent that universities distribute software, it is through nonexclusive copyright licensing.[71]

---

solutions and new tools to reach the community." Jason E. Stajich et al., *The Bioperl Toolkit: Perl Modules for the Life Sciences*, 12 GENOME RESEARCH 1611 (2002).

67 Interview with Brenner. *See also* Russ Altman et al., Whitepaper on Open Source Software in Bioinformatics (on file with author) (arguing that NIH should not mandate open source for its grant recipients)

68 *See supra* __ (empirical findings of non-restrictive licensing where audience is specialized).

69 *See* Rai and Sampat, supra note __ (discussing policies of 20 research universities with large software patent, biomedical patent, and/or biomedical research portfolios).

70 *Id.*

71 University representations are borne out by the data (to the extent the latter are available). One estimate based on using International Patent Classifications ("IPCs") common for software publishing industry patents indicates that universities are actually patenting less software as a relative matter in 2000 than they did in 1980. While university software patents represented 1% of all software in 1980, they represented 0.6% of all such patents in 2000. *Id.* Although IPC classifications, particularly those based on

Because of the relatively absence of TTO licensing activity in software, many universities are only beginning to formulate policies with respect to open source licensing.[72] Two universities that have relatively well-developed policies, the University of Washington and Georgia State, will treat software differently depending on whether it is perceived as commercially valuable.[73] For software that is not commercially valuable, the researcher's licensing preference will govern. If software is commercially valuable, both universities will recommend that software and source code be licensed free of charge to non-commercial users but licensed for a fee to commercial users. This differentiation between commercial and non-commercial can, of course, only be maintained through limits on redistribution of source code. Such limits are in tension with open source principles that counsel against impediments to redistribution. Limits on redistribution may also encounter research resistance: such resistance may stem not only from ideological commitments to open source but also from monetary incentives. To the extent that the software is commercially valuable, the researcher's financial interest may lie in widespread distribution of the source code to all potential customers. Such widespread distribution could be the best mechanism for maximizing consulting revenue. Unlike licensing revenue, consulting revenue does not have to be shared with the university.

A few universities do report "bright-line" policies regarding open source software that appear more encouraging to open source. For example, both MIT and Stanford allow different types of open source software licensing if the researcher wants to use that

---

publishing industry patterns, are admittedly a flawed metric they do give some sense of trends in university software patenting.

[72] *Id.*

[73] *Id.*

approach.[74] MIT also manages open source licenses for researchers. Similarly, the University of Texas defers in significant part to the licensing preferences of the researcher and also manages the researcher's licenses.[75]

2. Biological Database Projects

The first, and probably still most important, open and collaborative genomic database project was the publicly funded project to sequence the human genome. Unlike traditional human genetics, which revolved around individual laboratories that tended to be highly competitive – and hence, even prior to Bayh-Dole, were perceived as uneven in their willingness to share pre-publication information[76] – the group that formed the core of the Human Genome Project ("HGP") came from a less secretive community, worm genetics. The HGP was, from the outset, a collaborative endeavor. Not only did the sequencing laboratories work together but they all agreed at the outset to put their data into the public domain with in 24 hours.[77] But further articulation and codification of the task – in particular, the introduction of automated laser sequencing machines – made the collaboration run more smoothly. The intensity of the collaboration arguably increased further in 1998, after the project was faced with a challenge from Craig Venter, the leader of an effort by the private firm Celera to sequence the genome. After this challenge arose, major sequencing centers – the so-called "G-5" – were required to report their

[74] Interviews with Lita Nelsen, MIT, and Kathy Ku, Stanford

[75] Interview with Georgia Harper

[76] Interview with Huntington Willard; Walsh and Hong

[77] In January 2003, NHGRI extended this policy of immediate data deposition without accompanying intellectual property rights to all large-scale data "infrastructure" projects. Indeed, at this meeting, NHGRI prioritized immediate and full access to data over the traditional scientific norm that the investigator who generates the data has the right to do the first analysis of this data. 421 NATURE 875 (2003)

progress on individual chromosomes in weekly conference calls with the funding entities, principally the NIH's National Human Genome Research Institute ("NHGRI").[78]

The producers of the human genome sequence did not simply put the raw data into public domain. Rather, as the data were being produced, an open source software program known as the distributed annotation system ("DAS"), was set up to facilitate collaborative improvement and annotation of the genome. DAS has also been applied to other genomes, including mouse, C. elegans, fruit fly, and rice. Under the DAS system, any interested party can set up an annotation server. DAS enables end users of the information – in other words, researchers – to choose the annotations they want to view by typing in the URLs of the appropriate servers. Annotation quality is judged via consensus-based mechanisms. Specifically, according to Lincoln Stein, one of the designers of the DAS, it was "designed to facilitate comparisons of annotations among several groups. The idea is that an annotation that is similar among multiple groups will be more reliable than an annotation that is noted by one group."[79] The quality of the annotation is also judged by looking at published papers that describe the annotation technique.

Within the HGP, there was some discussion about using a type of copyleft license on the data produced by the project.[80] The view among these participants was that such a license would prevent private entities, particularly Craig Venter, from gaining an advantage over the public project by making proprietary any improvements Celera

78 John Sulston, The Common Thread.
79 Interview with Lincoln Stein.
80 JOHN SULSTON, THE COMMON THREAD 211 (2002)

made to the public data. Although the HGP leaders rejected a copyleft approach,[81] NHGRI, together with other funding organizations, has quite explicitly adopted a copyleft-style policy in setting up the International Haplotype Mapping Project ("HapMap"). This project aims to catalog haplotypes – patterns of genetic variation – and and link such patterns to disease phenotypes. In order to identify a particular haplotype, researchers must first identify the individual genotypic variations that make up the haplotype. The HapMap project is releasing individual genotype data as soon as it is identified. Before haplotype information has been assembled, it may be possible for those who access the data to take this data, combine it with their own genotype data, and generate enough information to file patent applications on haplotypes of interest. To address this possibility, the project has set up a click-wrap license that requires those who access the HapMap database to agree that they will not file product patent applications in cases where they have relied in part on HapMap data.[82] Although this license does not (and can not) rely on an assertion of copyright in the underlying data, it does represent an enforceable contract.[83]

Notably, with respect to all of these database projects, data dissemination and improvement policies have been developed by scientists and NIH administrators and essentially imposed on the administrators of the participating universities. Although universities have not played any role in formulating the policy, they appear to have

---

[81] Similarly, the participants in another important open and collaborative project that took place at approximately the same time as the HGP, the Single Nucleotide Polymorphism ("SNP") Consortium – which included 11 pharmaceutical companies and one non-profit partner, the Wellcome Trust – put its data in the public domain. For further discussion of the SNP project, *see infra* __.

[82] *See* International HapMap Project Public Access License, available at www.hapmap.org/cgi-perl/registration.

[83] Of course, because there is no underlying copyright, those who manage to access the data without having agreed to the license are not subject to any legal prohibition against patenting. The relative weakness of the HapMap prohibition is probably salutary, however, because, as discussed below, copyleft-style licensing for biological databases may impede commercialization unduly.

acquiesced in the rejection of proprietary rights.[84] Thus NIH has not needed to invoke the cumbersome legal procedure set up by Bayh-Dole to restrain university patenting.[85]

3. Wet Lab Systems Biology Projects

Outside the context of digital information produced through standardized protocols and machines – that is, for projects that require significant wet lab biology that is more difficult to divide into modules – the open and collaborative model has not been used as widely. However, it may be making some inroads in the context of some recent systems biology projects funded by NIH. In the last five years, the National Institute of General Medical Science ("NIGMS") has funded five large grants that are intended to "make resources available for independently funded scientists to form research teams to solve a complex biological problem that is of central importance to biomedical science . . . and that would be beyond the means of any one research group." These grants depart from the traditional biological grant model, which focuses on individual laboratories.

The Alliance for Cell Signaling ("AFCS") was the first of these large grants to be funded. AFCS's public funding is supplemented in part by funding from several large pharmaceutical firms. AFCS was inspired by the HGP,[86] and it clearly invokes significant elements of an open and collaborative approach. The Alliance is led by Nobelist Alfred Gilman of the University of Texas, Southwestern Medical School. Gilman won his Nobel Prize for his work on the role of G proteins in cell signaling, and the goal of the project is to map complex signaling networks. While cell biologists once believed that signals, such as a drug candidate binding to a cell receptor, initiated only

---

84 *See* Eliot Marshall, *Genome Researchers Take the Pledge: Data Sharing*, SCIENCE, April 26, 1996, at 478 (noting that key university patent officials approved of policy). One leading officer, Lita Nelsen of MIT, has noted, however, that she is wary of the "bad precedent" that the policy might set. *Id.*

85 See Arti Rai and Rebecca Eisenberg, Bayh-Dole Reform and the Progress of Biomedicine

86 Interview with Al Gilman

one pathway, it is now clear that a single chemical stimulus can excite different networks that interact in complex ways. As a consequence, promising drugs can have unexpected side-effects that cause them to fail in clinical trials. Combinations of ligands, which will be necessary to treat diseases influenced by multiple genes can increase complexity even further. The ultimate goal of the experimental work is to codify the "vast uncharted territory"[87] by generating a computational model of signaling within the cell.[88]

AFCS comprises eight "wet labs" and one bioinformatics lab. Each wet lab measures a distinct aspect of the effect produced by different ligands. The bioinformatics laboratory is responsible for integrating the data produced by the eight wet labs. The leaders of AFCS have determined that in order to generate reliable output that can be meaningfully compared, and aggregated, across labs, laboratory inputs (e.g. cell lines) and procedures must be standardized. Much work has gone into such standardization, and the protocols used are publicly available on the Web.[89] But standardization can only produce partial modularity. For the most part, the collaboration must employ a strategy in which each future experimental step is determined collaboratively, based on the data that emerged from the previous experimental step.[90] As a consequence, the AFCS laboratories must be in constant communication, both through videoconferencing and face-to-face meetings.[91]

Another novel aspect of AFCS involves its lack of emphasis, at least thus far, on conventional publication through peer-reviewed, printed scientific journals. Rather, after

87 *Id.*
88 Interview with lead bioinformatician Shankar Subramanian.
89 See www.signaling-gateway.org/data/Protocol/Links.html
90 Interview with Alex Brown, 4/21
91 *Id.*

some internal review, data publication takes place expeditiously on the Web.[92] Moreover, although AFCS investigators do replicate experiments and analyze data further – and publish those reviews on the Web as "research reports"[93] – they have no "head start" in terms of this analysis. In this respect, AFCS is explicitly modeled on the HGP. The lack of emphasis on conventional publication also coheres with the organizational structure of AFCS. While most lab directors are senior tenured professors who have advanced through the conventional career track for academic scientists, many of the individuals who work in AFCS laboratories are on a different, and in some respects novel, career track: they are postdoctoral scientists who are not planning on tenure-track appointments. Many of these individuals plan to go into private sector research.

On the other hand, AFCS is beginning to shift towards a more conventional, print publication–oriented approach. Some of the lab heads observe that it may be difficult to get scientists outside AFCS to pay attention to the data generated by the Alliance without using the conventional publication route.[94] Indeed, the AFCS web site now emphasizes that scientists who use AFCS data can publish their work in peer-reviewed publications; a few have in fact published such work.[95] Nonetheless, there is lingering concern that prestigious print publication venues – *Science*, *Nature*, *Cell*, and the like – may be reluctant to publish papers associated with data that is available on the Web prior to publication.[96]

---

92 See www.signaling-gateway.org/data/Data.html (AFCS data center, hosted by AFCS and Nature)

93 See www.signaling-gateway.org/reports/ReportCover.html

94 Interview with Alex Brown. Brown notes that most data, including the data generated by AFCS, is not as publicly visible as was the data from the Human Genome Project

95 See www.signaling-gateway.org/reports/JournalPubs.htm

96 Interview with Alex Brown.

Finally, and perhaps most unusually, all participants in AFCS have agreed to disavow intellectual property rights in their research. This agreement to disavow conventional property rights is, quite obviously, contrary to the trends in patenting that we have witnessed since passage of the Bayh-Dole Act. Moreover, many of the institutions participating in AFCS – perhaps most notably the University of California system but also the University of Texas and the California Institute of Technology – have substantial numbers of patents. It appears that Gilman's Nobel Prize, as well as his stature in the area of cell signaling, has enabled him to convince recalcitrant university administrators, particularly at the University of California and the California Institute of Technology, not to interfere in the disavowal of property rights. But even someone of Gilman's stature found the task difficult.[97]

Part III: Open and Collaborative Biomedical Research: A Critical Evaluation

This section assesses the extent to which the open and collaborative approach has the potential to produce socially desirable innovation, particularly as compared with more traditional, proprietary approaches. I also discuss a number of institutional obstacles to the adoption of this non-traditional model.

A. Open Source Bioinformatics Software

The variables involved in the normative evaluation of open source bioinformatics software are, to some extent, similar to those involved in the normative evaluation of open source software generally. Although no technical evaluation of which I am aware has specifically compared open source bioinformatics software with comparable closed source software, the technical superiority of certain types of open source software suggests that open source might, at least in some circumstances, yield technically

[97] Interview with Al Gilman

superior bioinformatics software. At a minimum, open source software may be a good alternative for producing an output of reasonable quality at low cost.

Additionally, because bioinformatics software development is generally done by publicly funded academics, incentives should not be a problem, even in the typical case where the contribution made by a particular academic is large. In addition to intrinsic incentives, the strong extrinsic incentive of achieving scientific kudos through publication is at work. As noted in the previous section, most open source bioinformatics software is developed to pursue a biological problem. Thus the public availability of information of the software in no way impedes publication in a conventional print journal of biological insights gained using the software.

It also appears unlikely that the type of specialized, small-scale software that is currently being developed by the open source bioinformatics community would in fact be developed by a commercially-oriented software firm. So there should be little fear that public funding is displacing production that would have taken place in the private sector in any event.[98] Moreover, if the open source software turned out to be valuable, commercial firms could presumably be built around providing services for the software.

The main obstacle to experimentation with open source bioinformatics software will arise in the small number of cases where such software is commercially valuable. As currently constituted, the financial interests of open source developers and university TTOs may be at odds. Universities can earn copyright licensing revenues only by restriction source code distribution to some degree. To the extent that open source developers derive money from consulting revenues, they may be reluctant to embrace any

[98] It is also unlikely that the "small firm" rationale for having upstream patenting, *see supra* __, applies. As contrasted with small biotechnology firms, small software firms do not appear to depends on upstream patents.

university restrictions on the availability of source code to potential customers. Indeed, in the last few years, there have been several celebrated cases where open source bioinformatics software developers have clashed with universities over questions of ownership and licensing. These have all been cases in which the researcher was deriving substantial consulting revenue from open source distribution, none of which had to be shared with the university.[99]

The question of who should be responsible for determining whether a particular piece of bioinformatics software is open source is a difficult one. Nonetheless, an argument can be made in favor of the approach taken by MIT, Stanford, and the University of Texas – deference to researcher choice. Although open source is not necessarily the best approach for all software projects (and thus, for example, any blanket NIH mandate to require open source software in its grants would be unfortunate), bioinformatics software developers are probably well-placed to determine whether, in any given case, open source development is the best approach as a scientific matter. Moreover, in the context of software, researcher choices are unlikely to impose significant development-impeding externalities (as they arguably did in the pre-Bayh-Dole era): the software itself can be developed through volunteer labor, and it is difficult to imagine how any form of open source licensing, even copyleft licensing, would undermine important patent rights on biochemical compounds, such as genes, proteins, or small molecule chemical drugs. Deference to researchers might be particularly desirable to the extent that researcher preferences were not unduly biased in favor of open source because of the prospect of consulting revenues. For example, universities might ask for the same percentage of consulting revenues that they currently get of licensing revenues.

[99] Memorandum from Pat Jones discussing cases.

This option would, of course, have the corresponding advantage of not biasing universities against open source.[100]

B. Biological Databases

As the leaders of the Human Genome Project often noted, particularly after the challenge from Craig Venture, an HGP-style approach to database generation provides an "infrastructure" of freely available scientific information for all researchers.[101] To the extent that property rights over such information would create inefficiently high transaction costs for follow-on wet lab innovation, free infrastructure is desirable. Indeed, the private sector – specifically the pharmaceutical sector – has on various occasions itself recognized this lesson. As noted earlier, pharmaceutical companies such as Merck have funded university researchers to place genomic information in the public domain. More recently, a consortium of ten pharmaceutical companies funded university researchers to find millions of single-base variations in the human genome and put them into the public domain. Pharmaceutical companies are likely to need information about many different single-base variations in order to tailor drugs to individual genotypes. Rather than deal with licensing a patent on each variation, these firms simply decided to preempt the possibility of patents.[102]

The public availability of biological databases is, of course, also necessary for distributed annotation. Like open source software development, DAS reduces the

---

[100] Another concern that has recently emerged is the possibility of contributors to open source adding code that may have property rights attached to it. Because it is very difficult for open source project leaders to verify that contributors are adding code that is free of proprietary rights, the SCO v. IBM lawsuit, in which SCO claims copyright interests over parts of UNIX that have allegedly been incorporated into Linux, has generated much concern in the open source community. At this early stage, the potential implications of this lawsuit, particularly for university-based open source researchers, are difficult to gauge, however.

[101] No doubt the analogy to publicly funded highways was not lost on the HGP leaders.

[102] The SNP Consortium. www.tsc.com

transaction costs associated with improving the information product. The result may be an information product that is ultimately superior to its closed counterpart.

Public databases differ from bioinformatics software, however, in that the high cost associated with generating initial data probably does undermine the ability to run a private database business on a services model. The race to sequence the human genome provided something of a natural experiment in this regard. Once the public data were available, the only additional value that the private Celera could provide was service-related. Although a significant number of firms and academic institutions did subscribe to the Celera database for these services,[103] the availability of the public data placed a ceiling on what Celera could charge. This ceiling was sufficiently low that Celera has largely moved out the database business and into drug development. To the extent that the challenge from Celera failed, we are likely to see fewer such challenges in the future.

For several reasons, this result could be seen as unfortunate. Arguably, the challenge from Celera provided the competition necessary for the public project to work more efficiently. Additionally, Celera made legitimate what was once considered a radical "shotgun" approach to genome sequencing. Had it not been for Celera's challenge, this approach, which is significantly cheaper and faster than the approach initially used by the publicly funded project, might not have achieved legitimacy as quickly. Finally, the challenge from Celera does raise the obvious question of whether publicly funded production is necessary when (at least absent the public effort) a small firm like Celera would have been able to market its database as a research input.

To the extent that the public effort can be justified (and I believe that on balance it can), this justification rests on two foundations: first, transaction costs militate in favor of

[103] JAMES SHREEVE, GENOME WARS 368-69 (2004)

upstream data being publicly available, even if such availability does require public funding and does undermine private database companies. In addition, to the extent that distributed annotation is likely to produce an information product that is superior to any that could be produced through firms or markets, a public database is necessary for such annotation to proceed.

The case for using copyleft licensing on public data is, however, much weaker. As with software, such licensing limits the availability of proprietary rights on downstream improvements. But unlike software patents patents on drug candidates – and perhaps even downstream research that leads directly to a drug candidate – are unequivocally important in the biopharmaceutical industry. Moreover, while copyleft licensing might be useful for inducing participation in purely volunteer open source projects,[104] it should not be critical for inducing such participation in projects where the collaborators are publicly funded academics. A statement by prestigious print publications pledging not to discriminate against articles that analyze publicly available data would provide additional incentives: although data generators and annotators might not have any official head start in submitting those articles, their familiarity with the data would make them most likely to submit the first analyses. Additionally, at least in the long term, it might be appropriate for the biological community to give data generators and annotators publication-type credit for their work, even if the work is placed immediately on the Web and reviewed by peers subsequent to such Web publication rather than prior to it.[105]

C. Wet Lab Systems Biology

---

104 *See supra* __.

105 This model has been used successfully, for example, in the physics community.

Thus far, the open and collaborative model's application to wet lab biology has largely been limited to systems biology. Even this limited application is quite significant, however. Unlike digital information projects, which are modular and hence can be pursued through a variety of alternative institutional mechanisms (even if less well through some mechanisms than others), systems biology is not modular. This means that the most plausible alternative to open and collaborative production is probably the large firm. Moreover, as discussed further below, even large firms may be reluctant to pursue risky areas of inchoate research independently.

As with initial data generation, the capital costs associated with wet lab biology are sufficiently high that it will probably be inefficient for most wet lab collaborations to be open to all comers. Indeed, even with a limited number of players, public funding will be necessary. Nonetheless, this number of players will still be significantly larger than in traditional small-lab based biological science. In addition, although this has not yet happened, it is certainly possible that annotation of the data generated by a wet lab collaboration could invoke the DAS model and thus encompass a larger group.

For collaborative projects in inchoate areas, the gains that can accrue from disavowal of intellectual property rights are particularly significant. Unlike AFCS, universities and investigators involved in other large-scale collaborative projects funded by NIGMS have not similarly disavowed such rights. Without disavowal of intellectual property rights, concerns that information exchange will lead to public disclosure of proprietary information, and disputes over how to allocate patent rights that might arise in the future between a host of different potential university assignees, may create friction. The principal investigator of one consortium that has not disavowed proprietary rights,

Rick Horwitz of the Cell Migration Consortium, reports some dissatisfaction with the manner in which the relevant university TTOs in his Consortium have conducted their negotiations. He believes that TTO-imposed requirements whereby each university agrees to keep strictly confidential, and to refrain from commercializing, the proprietary information of other universities in the Consortium have "gotten in the way of the science."[106] Horwitz hopes that the relevant inter-university agreements will be renegotiated in the future. The possibility that better agreements will be produced in the future is not necessarily high, however. Within wet lab biomedical research, universities jealously guard their ability to commercialize proprietary information, particularly by turning it into patents. In order to turn proprietary information into patents, strict restrictions on dissemination are necessary: the relevant court decisions by the Federal Circuit hold that even limited public sharing of information can create patent-defeating prior art.[107]

Problems of ownership and exploitation also make collaborative research difficult. Before research in an inchoate area has even been done, it is difficult to know how rights should be assigned to collaborators. In addition, in the patent arena, the default rules of ownership are quite unattractive. Under the default rule, any inventor who contributes to a single patent claim is considered a full owner of the patent. Thus, in the case of a university collaboration, even a small contribution from a university researcher would make the university a co-owner. Patent doctrine also allows each owner to exploit fully the patent without permission from the other owners and without

---

[106] Interview with Rick Horwitz

[107] The Cooperative Research and Technology Enhancement Act of 2004 (CREATE), recently passed by the Senate, aims to encourage collaborations by reducing the likelihood that so-called secret prior art created by the collaboration will defeat patentability. But this law does not address public disclosure of prior art created by the collaboration.

any duty to account. The combination of rules is such that collaborators rightly see the default as a situation studiously to be avoided.

On the other hand, an arguable advantage of the small contribution rule is that it forces disclosure of information to vulnerable parties who might otherwise be exploited.[108] The advantage of the default exploitation rule is that it avoids hold-up problems for owners who seek to license or other exploit the patent. Alternative sets of default rules may well have disadvantages that are correspondingly large.[109] The difficulty of designing good default rules simply underscores the difficulty of contracting for innovation in inchoate areas.[110] An obvious alternative to intricate contracting might be the formation of a new entity to which all rights could be assigned. But determining equity shares in this new entity could also be problematic.

These difficulties with contract-based collaboration might suggest a large firm would be the best locus for inchoate innovation. Indeed, as discussed earlier, many 20th century economists made precisely that claim. However, there is reason to believe that the hierarchical structure of large firms, particularly large pharmaceutical firms, is not conducive to innovative research. Additionally, even a large firm may consider such research too risky to take on itself. To the extent there is movement within the pharmaceutical industry towards doing large-scale research (as opposed to funding such research, as in the case of AFCS), pharmaceutical companies are contemplating an *inter-firm* collaboration. An organization called the CEO Roundtable on Cancer is considering a proposal to create a research collaboration among a large number of companies for

---

[108] Ian Ayres, Default Rules

[109] For example, the copyright defaults, which aggressively limit the extent to which a contributor can be a co-owner, and also require assent by all owners before licensing, have the corresponding problem of failing to protect vulnerable parties and creating possible anticommons difficulties.

[110] Cf. Gary Libecap's (discussing contracting difficulties back by parties engaged in oil drilling)

purposes of making an "all-out effort against cancer." Various pharmaceutical firms are considering this collaboration between firms even though the obstacles related to allocation of intellectual property rights – not to mention antitrust concerns – are obviously quite considerable.[111] Given the difficulty of setting up such a collaboration, the research may be better produced as a public good, as it is being produced by AFCS.

Though the AFCS approach is exciting and appears to be more feasible than inter-firm collaboration, it also has drawbacks. Perhaps most obviously, though they may be easier to secure than agreements assigning such rights, agreements to disavow intellectual property rights are hardly easy to achieve. Only charismatic and superbly credentialed scientists like Al Gilman are likely to be able to secure such agreements. They will do so by convincing their scientific colleagues that they must sign on to a particular research project, whatever the political difficulties. Moreover, even individuals like Gilman probably have to be supported by the relevant funding agency.

The disavowal of intellectual property rights in systems biology also raises concerns about commercialization. It is certainly possible that some of the information that is placed in the public domain by projects like AFCS will be left to linger in an undeveloped state, as feared by the proponents of Bayh-Dole. However, the fact that pharmaceutical companies are funding some of AFCS research indicates that commercialization is unlikely to be altogether defeated. More generally, the evidence that public domain status for upstream information defeats commercialization is hardly solid. What public domain status for such information may do is undermine certain small biotechnology companies. Even here, however, it is possible that the information

[111] Susan Warner, Collaborators Against Cancer, The Scientist, June 30, 2003 (noting these obstacles)

produced will be sufficiently far from patentability that biotechnology firms will be able to improve upon this information and market it as a patented technology.

From the standpoint of incentives, if projects like AFCS are going to work, it is clear that either their publication model or the print publication emphasis of the biological sciences needs modification. As matters currently stand, the lack of emphasis on publication, coupled with the disavowal of patents, has meant that only AFCS lab heads are traditional academics. To the extent that systems biology projects will succeed only if they attract the most creative young minds, the failure to attract such researchers is worrisome. It is important, therefore, that AFCS is moving towards more conventional publication for its own investigators. It is also important that AFCS is explicitly encouraging other investigators to use its data as the basis for publication in peer-reviewed journals. As noted earlier, a statement by prestigious peer-reviewed publications making it clear they will not discriminate against articles based on data that has already been made publicly available would be useful. In the long term, a move in the biological sciences towards a model that emphasizes Web-based publication, with subsequent peer review, is also worth considering.

Conclusion

Approaches to biomedical research in which such research is generated and improved upon in an open, collaborative fashion represent a potentially valuable experiment. The intuitively obvious loci for such experimentation are software and databases. In the case of software, the major obstacle to successful experimentation could be removed by instituting contractual mechanisms to divide consulting revenues between investigators and universities. With respect to open and collaborative databases,

the argument is somewhat more equivocal. Nonetheless, when the data in question are upstream, a significant case can be made in favor of publicly funded and publicly available databases that can then be improved upon collaboratively. The case becomes weaker as the information being produced is downstream in the research path. Rather than using copyleft style licensing that undermines patents on downstream information, it may be advisable to attract collaborators by attempting to change biological science norms regarding publication. Such norm change would also improve the value of experimentation with large scale collaboration in wet lab systems biology. In particular, it would help such collaborations to attract promising young investigators. Although open and collaborative research represents a paradigm shift for wet lab biology, experimentation with such a paradigm shift might be necessary for solving the intractable biological problems that are currently impeding development of breakthrough drugs. Moreover, to the extent that large scale wet lab collaborations that disavow upstream intellectual property rights can find support pharmaceutical company support, they are unlikely to undermine critical patents.