EXHIBIT 3

DOES PATENT STRATEGY SHAPE THE LONG-RUN SUPPLY OF PUBLIC KNOWLEDGE? EVIDENCE FROM HUMAN GENETICS

KENNETH G. HUANG
Assistant Professor of Management
Singapore Management University
Lee Kong Chian School of Business
50 Stamford Road #05-01
Singapore, 178899
Tel: +(65) 6828-0525
Fax: +(65) 6828-0777
e-mail: kennethhuang@smu.edu.sg

FIONA E. MURRAY
Associate Professor of Management
Massachusetts Institute of Technology
Sloan School of Management
50 Memorial Drive E52-567
Cambridge, MA 02142
Tel: (617) 258-0628
Fax: (617) 253-2660
e-mail: fmurray@mit.edu

September 2008

**Forthcoming Academy of Management Journal
Special Issue on Public Policy Implications of Management Research**

DOES PATENT STRATEGY SHAPE THE LONG-RUN SUPPLY OF PUBLIC KNOWLEDGE? EVIDENCE FROM HUMAN GENETICS

ABSTRACT

Knowledge-based firms seeking competitive advantage often draw on the public knowledge stream – ideas embedded in public commons institutions – as the foundation for private knowledge – ideas firms protect through private intellectual property (IP) institutions. However, we have a limited understanding of the converse relationship: the impact of private knowledge strategies on public knowledge production. We examine this question in human genetics, where policymakers debate expanding IP ownership over the human genome. Our econometric analysis shows that gene patents decrease public genetic knowledge, with broader patent scope, private-sector ownership, complexity of the patent landscape, and the gene's commercial relevance exacerbating their effect.

INTRODUCTION

How do firms' patent strategies, and the landscape of private property rights they collectively produce, influence the long-run production of public knowledge? Management scholars have paid close attention to the ways in which firms benefit from *public knowledge* – ideas disclosed through open commons institutions – by using it to generate *private knowledge,* protected by private property institutions such as patents (Cockburn & Henderson, 1998; Cohen & Levinthal, 1990; Fleming & Sorenson, 2004; Powell, Koput, & Smith-Doerr, 1996). However, they have paid scant attention to the converse relationship: the impact of private knowledge on public knowledge production. Instead, legal and policy analyses dominate the study of this relationship (Heller, 2008; Heller & Eisenberg, 1998; Lessig, 2004). This speaks to the importance of a management perspective bridging policy and legal studies with organizational theory and strategy, which can initiate a rich agenda examining the interaction between firm strategy and the institutional foundations of knowledge work.

The debate over public and private contributions to *genetic* knowledge exemplifies broader arguments over whether patenting helps or hinders public knowledge. It is fueled by the rapid sequencing of the human genome, the burgeoning stream of public genetic knowledge published in scientific articles and the expansion of gene patenting by industry. With patents protecting everything from full genes to small lengths of DNA (Andrews, 1991; Henry, Cho, Weaver, & Merz, 2002; Holman & Munzer, 2000), critics have argued that they increase gene test prices, initiate costly patent wars among firms and universities, and stifle public knowledge production in academia (Crichton, 2006; Eisenberg, 1996; Greenfield, 2006; Merz, 1999; Nash, 2000). Countering these claims, other scholars have noted that these patents provide important incentives for continued investment in useful genetic knowledge, drugs and gene diagnostics. A third, alternative perspective argues that private intellectual property (IP) rights on genes have no bearing at all on public knowledge in genetics (Walsh, Cho, & Cohen, 2005).

In seeking clear answers to the question of how gene patenting influences the long-run supply of public knowledge, scholars to date have relied on evidence from individual cases highlighting the aggressive enforcement of gene patent rights over academic scientists (Henry et al., 2002). While richly describing the challenges of gene patents, these studies cannot illuminate the extent of these practices across a wide variety of patents and genes. A better approach would be to rely on large-scale empirical studies. However, several inherent challenges account for why such studies have not been forthcoming until now. First, until the recent documentation of the patent landscape of the human genome (Jensen & Murray, 2005), systematic data on private (patented) genetic knowledge was limited. Second, even with such data, traditional approaches cannot estimate the causal impact of patenting on the public knowledge stream: given the possible variations in knowledge associated with patented and unpatented genes, simple comparisons are uninformative. A third issue further confounds the problem: confusion as to whether the public and private knowledge streams should be defined by different types of knowledge (basic v. applied), the organization of knowledge production (academia v. industry), or the institutional sphere defining knowledge disclosure, access and accumulation (public commons v. private property). Finally, management theory has no synthetic framework in which to analyze disparate evidence on the relationship between the public and private knowledge streams.

This paper aims to address these issues. It develops a framework highlighting institutional differences that arise when researchers embed knowledge in the public commons and in private property (Dasgupta & David, 1994; Weitzman, 1974), and explores the interrelationships between these spheres (Colyvas & Powell, 2006; Murray & O'Mahony, 2007). We recognize that, in many instances, the same "pieces" of knowledge are contributed to both public and private knowledge streams through its disclosure in both publications and patents – a feature referred to as patent-paper pairs (Ducor, 2000; Murray, 2002). More than simply an artifact of

knowledge disclosure, these pairs instantiate a dynamic process through which public domain knowledge is subject to informal norms (for exchange and follow-on use) until the paired patent is granted (an average of 3 to 4 years). With the grant of private property rights, the knowledge is subject to the formal legal rights associated with IP institutions. Consequently, while knowledge in the public and private streams can converge through patent-paper pairs, the normative requirements of the public and private institutions may diverge and, at times, conflict.

Patent-paper pairs allow us to assess the implications of making joint contributions to the public and private knowledge stream. They also make it possible to compare the pre- and post-patent periods closely and observe the ways in which the institutional foundations of private knowledge streams shape the long-run public knowledge stream (Murray, 2008). To estimate the effects of these institutional interactions, we extend novel econometric methods into the management domain. Using the differences-in-differences approach from economic analyses (Furman & Stern, 2006; Rysman & Simcoe, 2008), we assess the impact of patent grant on the rate of production of follow-on public knowledge by examining the annual number of papers citing the focal gene paper in the years before and after patent grant (Murray & Stern, 2007a). This allows us to develop the first systematic, large-scale quantitative evidence of the causal impact of gene patent in human genetics.

The starting point for our study is the population of 4,270 human gene patents (covering almost 20% of 23,688 known human genes) (Jensen & Murray, 2005), from which we identify 1,279 human gene patent-paper pairs. These pairs are distinguished by the shared disclosure of a gene sequence in the "gene paper" and in the claims of the "gene patent". The gene sequences are then traced to all other patents issued on the gene, from which we build measures of the patent landscape by devising novel metrics of patent ownership fragmentation (Ziedonis, 2004) and overlapping patents (Shapiro, 2001). We then investigate how individual patents, and the patent landscape they produce, together contour future public knowledge production.

This paper responds to Hitt's call (2005) to apply management theory and tools to enrich public policy. By doing so, we provide policymakers and executives with evidence to guide their decisions over gene patents and the patent system more broadly. These issues have the potential to shape the institutional foundations of U.S. innovation for years to come. A change in the scope of gene patents or an outright ban – as proposed in a 2007 Congressional Bill (H.R. 977, 110th Cong) – could transform private knowledge streams (and alter opportunities for competitive advantage) and indirectly shape the long-run public knowledge stream in sectors as diverse as human health, agriculture and the environment. Beyond human genetics and the life sciences, our findings help guide the study of patenting in other knowledge-intensive industries, especially the question of how patent strategies, and the patent landscape firms collectively produce, affect follow-on knowledge in the public and private spheres.

The conceptual framework at the heart of our analysis has important implications for management scholarship, an arena where scholars currently limit their research to the one-way relationship between public knowledge and its incorporation into private knowledge streams. Our more comprehensive perspective, capturing as it does the dynamics in both directions, may reveal subtle positive and negative impacts that would otherwise elude researchers. This approach is particularly salient as firms seek to engage strategically with knowledge communities producing a variety of public goods through voluntary contributions to the public domain – from scientific knowledge to software, designs and music (von Hippel & von Krogh, 2006). Finally, our analysis initiates further research at the intersection of three literatures: institutional theory, with its insights into the formal and informal norms of knowledge work (Colyvas & Powell, 2006; Dasgupta & David, 1994; Fauchart & von Hippel, 2008); intellectual property strategy (Gans & Stern, 2000; Teece, 1986; Ziedonis, 2004), which focuses on how firms use intellectual property to shape strategic outcomes; and lastly, organization theory, where

the emergence of organizations facilitating knowledge work takes center stage (Brown & Duguid, 2001; Hargadon & Bechky, 2006; O'Mahony & Bechky, 2008).

THEORY AND HYPOTHESES

Institutional Foundations of Private and Public Knowledge Streams

Scholars have long examined the complex dynamics of knowledge production and accumulation across firms, communities, and regions (Almeida, Dokko, & Rosenkopf, 2003; Hansen, 1999; Kogut & Almeida, 1999; Kogut & Zander, 1992). A central feature of these analyses is the recognition that streams of knowledge are embedded in distinctive institutionalized spheres – public and private – that shape the rules of knowledge disclosure, access and reward (Dasgupta & David, 1994; Murray & O'Mahony, 2007; Weitzman, 1974). It is widely assumed that we can map the institutional choice for any piece of knowledge from the type of knowledge and the organization of its production. In the canonical formulation, basic knowledge is generated in academia and then embedded in public commons institutions. Conversely, applied knowledge is generated in industry and embedded in private property institutions. While mounting evidence suggests that this simple mapping is no longer valid (Gans, Murray & Stern, 2008), distinctions between the two institutional spheres remain salient.

Public knowledge, embedded as it is in the public institutional system, is disclosed into the public commons and there exist few limitations to its follow-on use by others. The limited institutional norms shaping access are informal and governed by "Republic of Science" (as articulated by Merton (1973) with caveats acknowledging substantial local variations (Knorr-Cetina, 1999; Latour & Woolgar, 1979) or other informal norms (Fauchart & von Hippel, 2008). Access (in the case of scientific knowledge) requires acknowledgement through citations, rewarded via later recognition (de Solla Price, 1965; Hagstrom, 1965). As each piece of knowledge disclosed through a publication contributes to a public knowledge stream – the flow

of public knowledge –it provides the foundation for the future knowledge with the inter-temporal contributions to the public knowledge streams linked through citations (Garfield, 1955).

In contrast, the private knowledge stream is produced when knowledge is disclosed in patents and embedded in an intellectual property sphere. In return for disclosure, patent owners receive a time-limited monopoly over their knowledge, which enables researchers to prevent others from using their knowledge or to insist that follow-on innovators secure a license and make a variety of payments, including royalty payments or fees. In parallel with the normative citation requirements linking papers in the public knowledge stream, patent law incorporates an enforceable obligation to cite prior patents when an innovator builds on prior works. Therefore, in empirical terms, the citation link from one patent to another provides the (albeit limited) trace of the private knowledge stream, linking one generation of knowledge production to the next (Trajtenberg, Henderson, & Jaffe, 1997).

Interrelationships between Public and Private Knowledge Streams

Understanding the interrelationships between the public and private knowledge streams – and their corresponding institutional spheres – can provide important insights into the flow of knowledge throughout the economy, and more specifically, the ways in which firms can manage and capitalize on these streams. To date, however, scholars have not integrated these complex relationships into a complete theoretical framework. This synthesis has been elusive, in part, because the type of knowledge (basic v. applied), the locus of its production (academia v. industry) and the institutional sphere where it is embedded (public v. private) have been confounded or have only been considered in fragmentary ways in disparate literatures. For example, the analysis of firm-level patenting and publishing also does not address whether these two forms of disclosure capture the same or different types of knowledge (Gittelman & Kogut,

2002; Lim, 2000). Our goal then, is to synthesize these disparate literatures as a step toward a deeper understanding of the interrelationship between public and private knowledge streams.

Public Knowledge Stream Influences on the Private Knowledge Stream. Traditional perspectives on competitiveness and long-term economic growth underscore the central role played by the public knowledge stream as the foundation of the private (patented) knowledge stream (Romer, 1994). At the highest level of abstraction, this relationship is described in terms of the linear model of science: advancements in the public knowledge stream are equated with progress in basic science, which in turn establishes critical inputs for the downstream private (patented) stream of applied research, technological innovation and commercialization (Rosenberg, 1974). Vannevar Bush famously articulated this view in his call for heightened funding of the "endless frontier" of public scientific knowledge (Bush, 1945).

Management theorists often echo this view when they examine specific mechanisms firms use to access and leverage public knowledge streams. The notion of absorptive capacity (Cohen & Levinthal, 1990) captures the possibility that firms can, and should, absorb knowledge available in the public commons as part of their attempts to make effective contributions to the private knowledge stream. Moreover, it is argued that firms establishing strong capabilities in absorbing the public knowledge stream will be more effective innovators and therefore more competitive (e.g. Powell et al., 1996; Zucker, Darby, & Brewer, 1998). However, scholars disagree not only on how to measure this linkage,[1] but also on the specific ways in which firms engage with the public knowledge stream. There is also little consensus on the impact of those interactions on firms' contributions to private knowledge streams. Some, for example, argue that these public-private linkages are a rich source for new scientific techniques (Klevorick, Levin, Nelson, & Winter, 1995) contributing to applied knowledge. On the other hand, engagement

---

[1] At least three measures are used to capture the relationship between public and private knowledge streams: The number of publications cited in patents, or "science linkage" (Narin, Hamilton, & Olivastro, 1997; Tijssen, 2002), the patent and publication portfolio of firms (Gittelman & Kogut, 2003; Lim, 2000), and co-authorship and co-patenting networks (Owen-Smith & Powell, 2003; Powell et al., 1996; Zucker et al., 1998).

with the public knowledge stream may provide stronger intellectual foundations for private knowledge production (Nomaler & Verspagen, 2007). Specifically, building on Nelson's early formulation (1959), Fleming and Sorenson (2004) argue that firms use public knowledge streams as a "map" to help navigate and contribute to the complex landscape of patentable innovations.

A third interpretation suggests that engagement with the public knowledge stream provides industrial scientists with non-monetary rewards, access to conferences and academic scholars (Stern, 2004). Moreover, by pursuing a strategy that rewards contributions to knowledge embedded in and endorsed by the public institutional sphere, firms might be able to attract and retain higher quality researchers (Henderson & Cockburn, 1994; Cockburn & Henderson, 1998).

Private Knowledge Stream Influences on the Public Knowledge Stream. Management scholars have devoted less attention to the ways in which decisions to participate in the private knowledge stream may influence contributions to the public knowledge stream. Studies of industry's influence over the direction of public knowledge production hint at a linkage from private to public knowledge, forming a feedback loop that reverses the traditional linear model of science (Meyer, 2000; Rosenberg, 1974). However, this formulation focuses on the content of the public and private knowledge streams and not on whether the private influence on the public knowledge is grounded in corporate decisions over IP rights.

Do such institutional decisions positively or negatively influence future contributions to the public knowledge stream? The literature posits a number of possible mechanisms of influence. At the broadest level, some question the morality of knowledge production undertaken (by academics) in the shadow of private (intellectual property) interests. They argue that private property rights and their related financial interests undermine academic objectivity, causing bias, suppression of results, and even fraud (Krimsky, 2003; Resnik, 1998a, 1998b). More concretely, they claim that contributing to private knowledge streams, in the form of patents, will encourage

secrecy (in the form of timing decisions, withholding of information, materials etc.) and lower contributions to the public knowledge stream (Blumenthal, 1997; Campbell et al., 2002).

Other scholars have focused on how the rewards of patenting distort researchers' research agendas (Thursby & Thursby, 2002). In this view, researchers will shift towards more applied research in order to contribute more effectively to the private knowledge stream (Aghion, Dewatripont & Stein, 2005). This parallels the countervailing concern that firms contributing to the public knowledge stream weaken their ability to generate private knowledge (Gittelman & Kogut, 2003). Many of these worries about public-private influence are premised on the idea that research projects conducted in public and private organizations differ. However, evidence suggests that this assumption is false: scientists engaged in a range of endeavors can contribute knowledge to either or both knowledge streams – from early semi-conductors to recombinant DNA and software code (Murray, 2002, 2008). In other words, the content of the two knowledge streams can converge even while their institutional spheres remain distinctive.

Convergence of the Private and Public Knowledge Streams

Researchers in universities and firms confront the possibility that their research results can contribute to either the public or the private knowledge streams. Specifically, they can choose to embed their knowledge in both institutional spheres in the form of a patent, a publication or both – a patent-paper pair. These pairs are distinctive instantiations of the same pieces of knowledge whereby the knowledge disclosed in the paper is also the subject of patent claims (Ducor, 2000; Murray, 2002). Far from being an unusual or strikingly modern occurrence, researchers from Pasteur to Shockley and later Cohen and Boyer have disclosed their knowledge in patent-paper pairs, thus contributing to the convergence of the public and private knowledge streams. For example, while doing research at Bell Labs, William Shockley created the foundations for early semiconductors with his development of the transistor. The experiments he undertook in January

1948 are described in his Bell Labs lab notebook (No. 20455 pp: 128-32 (January 1948). Less than six months later, in June 1948, he filed for a US patent on the solid-state transistor (*U. S. Patent 2,569,347* issued September 25, 1951). In 1949, he published the theory underlying the transistor inventor (Shockley, 1949). Pairs are also prevalent for knowledge in human genetics (Huang, 2006; Huang & Murray, 2008), a finding that lies at the heart of our analysis.

More than simply a curious object and useful methodological device, contributions to the public and private knowledge streams in the form of patent-paper pairs initiate complex dynamics and potential collisions between the distinctive normative requirements of the public and private institutional spheres (Murray, 2008). After generating new knowledge, researchers can choose to document their findings in the public domain, either directly or more commonly via traditional peer-reviewed journals. In many scientific fields, manuscripts move through review to publication in three to six months, thus rapidly joining the public knowledge stream. With publication, researchers are now subject to the informal norms of public science; material sharing, data exchange, assistance with tacit knowledge etc. (Kohler, 1994). While fields vary in the specifics, in most instances, exchange of information, materials, methods, and other knowledge required to replicate or accumulate is accompanied by an obligation to acknowledge, cite or possibly to collaborate. Beyond these expectations, there are few strictures on follow-on researchers as they draw from and build on the public knowledge stream.

These limited obligations remain salient even if researchers also submit their manuscripts to patent lawyers (directly or via a technology transfer office (TTO)) in order to file patents. They face one key constraint: the patent must be filed within a year of the paired publication (or disclosure of the paired knowledge at a conference) in order to retain U.S. patent rights (foreign private property rights are lost if the patent application is filed after publication). Once researchers have met this constraint, the public and private regimes do not intersect again until the patent is granted – typically three to four years after the paired publication. In this period, the

norms of the public commons apply to researchers. However, in the post-patent grant period, researchers and their firms have formal legal IP rights over the knowledge disclosed in the patent and in the paired publication. Thus, while still bound by the norms of the public commons, legal IP rights can be invoked. In particular, researchers may provide, restrict or prohibit access to those seeking to draw from and build upon their contribution to the private knowledge stream for follow-on knowledge production.[2] The timeline described above for the patent-paper pair is illustrated in Figure 1.

```
----------------------------------
        Insert Figure 1 about here
----------------------------------
```

Even academic researchers who do not read the patent (Walsh et al., 2005), but who use the knowledge in the paper, now fall within the scope of possible IP enforcement as clarified in the *Madey v. Duke University* decision (Dreyfus, 2004). Not limited to industry-owned patents, the University of Wisconsin has also imposed its rights to human embryonic stem cells over researchers in academia and industry through complex licensing agreements (Murray, 2007). Consequently, while the actual knowledge in the public and private streams converges through patent-paper pairs, the normative requirements of the two streams do diverge and may conflict.

What is the actual impact of placing knowledge in both the public and the private knowledge streams? How do these conflicting regimes intersect, and how are the tensions resolved? Addressing these questions requires a more systematic analysis of the ways in which patent strategies of firms affect the public knowledge stream. If patents stifle follow-on knowledge contributions to the public stream, this raises the possibility that a firm's patent strategy could undermine the long-run supply of ideas to the public knowledge stream. Prior to the analysis we report here, empirical studies have not fully documented whether the expansion of patent rights,

---

[2] Exceptions to the "refusal to license" arise when the Federal government exercises 'march-in' rights (insisting on licensing ideas they funded), and when antitrust is binding (Lewis & Yao, 1995, MacKie-Mason, 2002)

and the complex patent landscape produced by the combined patent strategies of multiple firms and universities, have had a causal impact on subsequent contributions to public knowledge. We ground our predictions and analysis in the observation of patent-paper pairs, specifically the impact of patent grant on the future public knowledge stream that builds upon the paired paper.

Theoretical arguments highlight the rising numbers of patents over inputs into scientific research (such as gene sequences, cloned laboratory animals, and research reagents), and argue that these patents increase experimental costs and administrative burdens as scientists seek access to a wide variety of inputs, resulting in lower production of public knowledge – an anti-commons effect (Heller & Eisenberg, 1998). Published empirical evidence supporting this claim is conflicting. On the one hand, surveys suggest that academics ignore patent databases during research planning and that the most arduous restrictions come from material transfer (Walsh, Arora, & Cohen, 2003; Walsh et al., 2005). On the other hand, recent quantitative studies of a sample of patent-paper pairs (based on a sample of *Nature Biotechnology* papers) show a modest (10%) decline in the number of times the paired publication is cited in future publications after the corresponding patent is granted (Murray & Stern, 2007a). This decline was strongest in the late 1990s and early 2000s as researchers struggled to adapt to new legal constraints (Murray & Stern, 2008). We therefore hypothesize:

Hypothesis 1. The grant of a patent over knowledge claimed in a publication will negatively impact the rate of follow-on publications.

Most firms follow complex patent strategies to strengthen their IP rights by expanding their breadth and strength. Scotchmer (1991) argues that broader patents will more likely deter follow-on contributions to the private knowledge stream by providing stronger disincentives for follow-on knowledge production. Following this reasoning, Lerner (1994) shows that patents with broad scope – measured by the number of patent classes – are more valuable. Moreover,

they will likely impinge upon many different follow-on research lines and thus have a deterrent effect on a larger number of follow-on researchers (Aghion et al., 2005). By the same logic, we argue that broader patents will be more likely to stifle follow-on knowledge production contributing to the public knowledge stream. Thus, we hypothesize:

Hypothesis 2. The broader the patent granted over the knowledge claimed in a publication, the more it will negatively impact the rate of follow-on publication.

The number of claims captures patent strength: they are the legal articulation of the boundary of the property rights, with the principal claim defining the essential novel features of an innovation. As such, claims measure the extent of the innovation protected by the patent (Harhoff & Reitzig, 2004; Lanjouw & Schankerman, 2001). Recent evidence shows that the number of patent claims is a key determinant of the value of patent licensing agreements (Gans, Hsu, & Stern, 2008). We therefore postulate that stronger patents, proxied by the number of claims, have a more negative impact on follow-on pubic knowledge streams.

Hypothesis 3. The stronger the patent granted over knowledge claimed in a publication, the more it will negatively impact the rate of follow-on publications.

The type of patent owner (assignee) is a critical determinant of patent enforcement strategies (Somaya, 2003). For-profit private corporations have a high propensity for patent acquisition, enforcement and litigation, particularly compared to their public sector counterparts (Ziedonis, 2004). Assuming that follow-on researchers planning to contribute to the public knowledge stream have similar expectations regarding the propensity to litigate, then patents assigned to private sector (versus those assigned to the public sector) will have a greater negative impact on the number of times a paired paper is cited in future papers. Therefore, we hypothesize:

Hypothesis 4: Grant of a patent over knowledge claimed in a publication will more negatively impact the rate of follow-on publications when the assignee is from the private sector compared to the public sector.

In some instances, a single patent-paper pair on a foundational piece of knowledge with few substitutes may reduce follow-on contributions to the knowledge stream. However, Heller and Eisenberg's argument (1998) is grounded in the patent landscape surrounding public knowledge, not a single patent: the complex of patents that collectively impinge upon follow-on research. Faced with this landscape, researchers must navigate a morass of interpenetrating patent rights in order to contribute to the public knowledge stream. As Shapiro (2001) describes: "what happens if, in order to scale the pyramid and place a new block on the top, a researcher must gain the permission of each person who previously placed a block in the pyramid, perhaps paying a royalty or tax to gain such permission?" In testing whether this effect shapes contributions to the public knowledge stream, we rely on two characteristics that define the patent landscape: thickets and fragmentation. A patent thicket is "an overlapping set of patent rights requiring those seeking to commercialize new technology to obtain licenses from multiple patentees (Shapiro 2001, p.1)". We measure this as the number of patents claiming the same set of (gene) knowledge inputs and hypothesize:

Hypothesis 5: The denser the thicket of patents over knowledge claimed in a publication, the more it will negatively impact the rate of follow-on publications.

While patent thickets may be salient to follow-on researchers, the fragmentation of these multiple competing patent rights across many owners is potentially more problematic for follow-on researchers. Heller and Eisenberg (1998) outline the rationale behind this assumption, arguing, "a proliferation of patents on individual fragments held by different owners seems inevitably to require costly future transactions to bundle licenses together... Such a proliferation

of claims presents a daunting bargaining challenge. Unable to secure a complete set of licenses, firms choose between diverting resources to less promising projects with fewer licensing obstacles or proceeding... on the basis of incomplete information" (p. 699). Measuring fragmentation by the number of owners in a patent thicket, we hypothesize:

Hypothesis 6. The greater the ownership fragmentation of patents over knowledge claimed in a publication, the more it will negatively impact the rate of follow-on publications.

We take the view that on balance, granting of paired patents has a negative effect on the rate at which follow-on knowledge accumulates (in subsequent publications). Not all publications are the same, however, and so our claims must account for the type of knowledge captured in the converging public-private knowledge streams. When knowledge is immediately applicable to valuable problems and hence has greater downstream commercialization potential, the negative impact of the patent should be greater than for knowledge in paired papers with more ambiguous value. Our reasoning is that patents are "probabilistic" property rights whose enforcement is fraught with uncertainty and variance (Lemley & Shapiro, 2005). Consequently, follow-on researchers will assume that patents over immediately useful knowledge are more likely to be enforced that those over less useful knowledge. Thus, we hypothesize:

Hypothesis 7. The more useful the knowledge in a patent granted over knowledge claimed in a publication, the more it will negatively impact the rate of follow-on publications.

GENE PATENTING AND THE PRODUCTION OF GENETIC KNOWLEDGE

Concerns about how patenting affects the public stream of knowledge are found in many sectors of society (Heller, 2008). Indeed, any organization that gains competitive advantage through proprietary (private) knowledge and heavily depends on public knowledge must understand this relationship. Genetics is a uniquely suitable field in which to examine these

issues, not least because of the tremendous wealth of data on individual genes, from GenBank, the United States Patent Office, the National Center for Biotechnology Information, and the Online Mendelian Inheritance of Man database. It is also a setting where researchers incorporate knowledge into public and private knowledge streams through publications and patents, allowing us to identify patent-paper pairs, and where we can distinguish the commercial potential and other characteristics of different pieces of genetic knowledge. Finally, human genetics is of critical importance to health and welfare, used by firms as the foundation for innovation for many applications, from medical and environmental to industrial and agricultural products.

The gene patenting controversy exploded in 1991 when the National Institutes of Health (NIH) announced its patent applications on 350 gene segments discovered during the Human Genome Project (HGP). The NIH Director clarified that this was a temporary measure to protect the Federal government's right to these genes and to prevent preemptive patenting. In 1992, James Watson, then head of the HGP, countered that the move was "lunacy" (The Independent, 2007). The decision brought entrepreneurs, pharmaceutical executives, investors, lawyers, ethicists and scientists together in contentious debate. The validity of gene patents was uncertain, the scope of their property rights debatable, the ethics of the patents unclear and their impact on business and science contested.

Even in 1991, gene patents were not new:[3] In 1982, the first "gene patent" had been issued to the University of California (Patent Number 4,363,877). What changed was the possibility of patenting small chunks of DNA whose function was not yet determined, with no link to a specific disease and with only speculative utility. "Traditional" gene patents of the 1980s

---

[3] In 1980, the U.S. Supreme Court ruled 5 to 4 in *Diamond v Chakrabarty* (447 U.S. 303, 1980) to grant General Electric a patent on a genetically modified organism finding that "anything under the sun that is made by man," is patentable subject matter. This massively expanded the landscape of patentable genomic knowledge, including genetic engineering techniques, engineered DNA or RNA, plasmids, transformed cells, tissue cultures, cell lines, hybridomas, plants, animals, antibodies, antigens, hormones, other proteins, vaccines, and diagnostics (Restaino & Takeuchi, 2006). It also paved the way for patents on the use of human genes (DNA sequences in purified or isolated form, their usage and often their chemical composition).

resulted from labor-intensive gene hunting, with scientists gradually honing in on a particular gene implicated in a disease. These patents clearly identified the gene, its function, the protein it coded for and the disease for which the protein might be used. In contrast, the NIH patents were poorly characterized. Working for the NIH in the early years of the HGP, Craig Venter had developed a new rapid sequencing technique that decoded many small chunks of DNA called expressed sequence tags (ESTs); this neither required nor provided a deeper understanding of segment's function and usefulness. The NIH patents on these ESTs contained hundreds of gene segments but disclosed only a narrow understanding of their utility. The knowledge was limited to markers for individual chromosomes, primers, and templates (Holman & Munzer, 2000).

Arguments over the NIH patents sparked a broader debate around the pros and cons of all gene patents. On the positive side, biotech industry veterans noted that strong patent rights on human genes of therapeutic importance encouraged extensive investments in new therapeutic protein drugs. They also felt that if gene segments were useful they should be patentable. They were afraid to undermine the important principle that patents provide incentives for private sector investment in research (Borson, 1995). Many executives countered that strong patent claims to different parts of a gene would hamper follow-on research. For example, while a company might be willing to pay a single licensing fee for access to a full gene that they could use for drug development, they were reluctant to sign multiple licenses on a plethora of gene segments that happened to map to a gene they discovered through long and costly R&D (Henry et al., 2002). Research scientists concurred that gene patents could be problematic. Some objected to the prospect that a corporation would end up "controlling a priceless resource" (Nash, 2000). Others were reluctant to engage in curiosity-driven research about a particular gene if their results might be controlled by a private corporation.[4] Many believed that this type

---

[4] This fear was well founded: in 2000 Human Genome Sciences (HGS) received U.S. Patent No. 6,025,154 on gene CCR5. Their patent claimed the gene sequence and asserted it was a G-coupled protein receptor binding cytokines.

of patenting would stifle the international collaboration for public knowledge production that characterized the HGP and instead would lead to a patent war among research institutions (Andrews, 1991).[5] Some worried that a complicated patent landscape of gene segments could limit the freedom of researchers to determine the functionality of genes. They were concerned that researchers might get embroiled in complex licensing to access different parts of the human genome, and that this would lead researchers to shy away from working on the most interesting and relevant diseases. Public interest groups raised concerns that gene patents could increase the price of genetic tests to diagnose life-threatening diseases or limit the amount of medical research on such tests (Merz, 1999).

In October 1992, Human Health Secretary Sullivan reversed the NIH policy. Nonetheless, companies and academic institutions rushed into the genome land grab. By October 1996, the US patent office (USPTO) announced a backlog of patents with over 500,000 gene sequences (human and other species) – a project that would take one examiner 200 years and US$35 million to complete (Abate, 2000)! By the end of 1997, the USPTO had issued over 400 human gene sequence patents (Jensen & Murray, 2005). Throughout this period, gene patents were highly uncertain, probabilistic in nature and subject to wide variations in scope. In 2001, the USPTO finalized its utility guidelines for gene patents with the requirement that a human DNA patent must be specific – identifying a specific gene target, a biological reaction involving a specific protein and a real world use (Duke Law & Technology Review, 2001). Nevertheless, by then a significant fraction of the human genome was subject to a complex set of gene patents, with genes covered by up to twenty patents (Jensen & Murray, 2005).

In February 2007, the Genomic Research and Accessibility Act (GRAA) was introduced into Congress "to prohibit the patenting of human genetic material". For those called upon to

---

In 1996, a year after the HGS patent application, NIH scientists reported that CCR5 bound HIV and could be used as an HIV treatment. However, the HGS patent could be used to exclude scientists from developing such a treatment.
[5] Their fear was realized in June 1992 when the British government's Medical Research Council filed its own patent applications for 1,000 partial human gene sequences (Veggeberg, 1992).

respond to the Act, the lack of systematic data defining the effects of gene patenting made it difficult to move the discussion onto a firmer empirical footing and away from a collection of anecdotes. While our analysis does not examine the positive impact of patenting as it supports the private knowledge stream and thus spurs investment in human health, it does speak to the potentially detrimental impact of the IP on the long-run supply of public genetic knowledge.

METHODS

Empirical Approach

We analyze the impact of gene patents on the rate at which scientists contribute to the follow-on stream of public knowledge building on the gene papers, relying on several methodological and econometric advances. First, we use publication citations to each gene paper (i.e. peer-reviewed publications citing the focal paper) as a proxy for follow-on public knowledge accumulation. While we would like to measure the rate of follow-on knowledge production embedded in the public stream directly, it is difficult to gather systematic information for a large sample of publications. Our citation-based approach follows a long literature using citations to trace the flow of ideas and their follow-on accumulation in later knowledge production (de Solla Price, 1965; Hall, Jaffe, & Trajtenberg, 2001; Posner, 2000). The use of publication citations is subject to several caveats; first, it does not capture accumulation of non-disclosed knowledge. To counter this claim, we argue that a shift to secrecy still represents a reduction in follow-on contributions to the public knowledge stream because the reduction in *disclosed* knowledge has implications for future scientists. Second, we assume that published genetic knowledge is the relevant measure rather than forward citations in patents. While citations in patents are also important, we are interested in whether there are changes in follow-on contributions to the *public* knowledge stream when patents claim the same knowledge. Third, and most problematic, is that scientific citations capture follow-on citing behavior rather than follow-on knowledge

production. There is some evidence that citations are influenced by considerations beyond an information signal of knowledge exchange including friendship (Stigler & Friedland, 1975) and gender (Ferber, 1988). However, there is no direct evidence suggesting that citation behavior is narrowly contoured to the precise timing of events such as patenting or patent enforcement.

Predicting the impact of gene patents on the follow-on public knowledge stream raises several identification challenges. Without a well-identified "experiment", it is difficult to disentangle multiple factors underlying observed changes in follow-on knowledge accumulation. For example, any observed difference in citations between gene papers associated with patented and unpatented genes are difficult to explain. First, gene papers associated with patents may simply be different in *quality* (higher or lower) compared to gene papers on unpatented genes. Indeed a growing body of evidence points out that scientific knowledge both published in academic journals and patented is, on average, more highly cited over its lifetime than unpatented articles (from the same journal) (Huang, 2006; Murray & Stern, 2007a). Second, genes associated with patents may be *different* from non-patented genes and those gene papers might thus exhibit different citations patterns. We deal with these issues by focusing only on gene papers associated with patented genes. We also take advantage of many observable measures of the genetic knowledge disclosed in a gene paper: the strength of its link to a particular disease, and the size of the current body of knowledge of its function etc. However, our central methodological approach is the development and analysis of a sample of gene patent-paper pairs. Adopting Murray and Stern's (2007a) identification strategy, we use patent grant as an exogenous shock and examine publication citation rates to the paper in the pre- and post-grant period (with the former serving as a control group for the latter). As noted above, in the period prior to patent grant, genetic knowledge is disclosed in a paper and contributions to future public knowledge stream building on that paper accumulate in an institutional setting characterized by public norms and practices. This environment changes with the grant of the paired gene patent,

and follow-on accumulation takes place in the shadow of formal legal private property. By comparing the difference in gene paper citations in the pre- and post-grant period for those affected by the patent grant to the same difference for unaffected gene paper citations, we evaluate the precise impact of the change – the differences-in-differences identification.

The management literature and the program evaluation literature in economics use a variety of similar changes in the institutional environment of a piece of knowledge as an empirical strategy to deepen our understanding of the innovation process. The impact of incorporating a patent into an industry standard is assessed by comparing the pre- and post-standard setting rate of patent citations (Rysman & Simcoe, 2008). Likewise, management scholars use the exit of a firm from a product line or sector to explore whether a firm's innovative knowledge dies with it, even though it is already disclosed in patents (Hoetker & Agarwal, 2007). More salient for our analysis, Furman and Stern (2006) explore whether depositing enabling materials linked to specific publications increases the cumulative public knowledge stream measured with citations.

The identification approach implicit in gene patent-paper pairs relies on two assumptions. First, because differences in the time between scientific publication and the granting of the paired patent provide important variation, we must determine whether the timing of gene patent grant is random, and assume that it is not anticipated by those who accumulate and cite the gene paper in the future public knowledge stream. This assumption is particularly strong in the gene patenting case given the significant uncertainty in the likelihood of granting a gene patent in our time period. Second, we assume that gene patent-paper pairs claim the same piece of knowledge. We do this by precisely linking the patent and the paper via the disclosed gene sequence. This improves on the specificity of prior works (e.g. Murray & Stern, 2007a; Sampat, 2005) by ensuring the patent and paper are "paired" by knowledge claimed in the patent.

Data, Sample and Measures

We construct a novel panel of gene patent-paper pairs based on the population of 4,270 U.S. patents claiming uses of human genes as identified by stringent bioinformatics criteria (Jensen & Murray, 2005). The definition of a gene patent is specific: the gene sequence must be at least 150 nucleotides in length and match (e-value of zero) a human gene sequence (mRNA transcript) rather than another organism in the National Center for Biotechnology Information (NCBI) RefSeq public database. The resulting U.S. patents covered 4,382 or 18.5% of all known human genes (approximately 23,688 genes). Figure 2 shows the distribution of these 4,270 granted gene patents by application year.

Insert Figure 2 about here

To develop and identify the patent-paper pairs for these gene patents, we used a semi-automated search of the Thomson ISI Web of Science database which offers comprehensive coverage of peer-reviewed scientific research articles: (1) all patent inventors must appear as authors (a more stringent matching algorithm than that followed by Murray and Stern (2007a)); (2) publication must include the disclosed gene sequence; (3) patent abstract and application dates must fit the publication abstract and publication dates.[6] The match produced 1,498 matched patents. Finally, in those instances where a gene paper was paired with more than one gene patents, we included only the first patent as the paired patent (although the subsequent patents were included when we constructed our patent thicket measures). The final sample consists of 1,279 unique patent-paper pairs covering 2,637 genes. To ascertain whether this paired patent sample differed from the full gene patent population, we compared the means and

---

[6] The authors hold degrees in biomedical engineering and applied chemistry. In almost all the cases, the patent-paper pair assignment was unambiguous.

standard deviations of observable patent characteristics of the final sample versus starting population.[7] We found almost all of them to be statistically similar.[8]

The variables used in this study are based on observable characteristics of the gene papers and their citations, gene patents, and genes sequences as described in Table 1. Table 2 provides descriptive statistics, while Table 3 shows the correlation matrix for the variables. We draw our data from five different sources. Data for the gene papers and citations are based on ISI Web of Science. Data for the gene patents is from the USPTO. Gene characteristics are gathered from NCBI, the Online Mendelian Inheritance in Man (OMIM) database of disease genes and a database of cancer genes – the Sanger Institute Cancer Gene Census.[9]

Insert Table 1 about here

Insert Table 2 about here

Insert Table 3 about here

Citation-year variables. *Annual cite,* our dependent variable, counts of the annual number of citations the gene paper receives beginning in the year the paper was published, continuing until 2006. As a conservative test, we report results that exclude any author self-citations – defined as citations of papers written by any author of the focal paper.[10] The total number of gene-paper year observations is 12,830. The mean level of *annual cite* is thirteen (with a

---

[7] These observable characteristics include patent application and grant year, patent grant lag, number of national classes, type of national classes, number of claims, number of inventors, number of assignees, number of cited patent references, number of citing patent references, number of non-patents cited, and several constructed patent measures based on Trajtenberg, Henderson and Jaffe (1997).

[8] Number of inventors, classes and non-patents cited differ slightly: 2.6 to 3.3, 6.2 to 6.3, and 479 to 459 in the sample versus population respectively. The actual differences in magnitude in all three cases are trivial.

[9] First published in Futreal, P.A. et al. (2004) "A Census of Human Cancer Genes." *Nature Reviews Cancer,* 4: 177-183, which summarizes more than two decades of searching. The Cancer Gene Census is updated on Wellcome Trust Sanger Institute (2006): http://www.sanger.ac.uk/genetics/CGP/Census/

[10] As an additional test we used two other variations of the citation data: (1) excluding organizational self-citations (defined as citations of papers written by any author from the same organization as the focal paper); and (2) including author and organization self-citations. In both cases, the results remain essentially unchanged: the directions of the coefficients are similar and the differences in their magnitudes are very small.

minimum of zero and a maximum of 294) and by the end of the period, the average article has accumulated 148 citations over its lifetime as measured by the *total cite* for each gene paper. The *citation year* measures the calendar year in which a given citation is made. The *paper age* describes the age of the gene paper when a given citation is made, thus a citation made in 2000 to a paper published in 1998 is two years old. The average age of a citation in the sample is about five years, while the average *citation year* is 2001.

Paper variables. Several independent variables capture characteristics of the gene paper. *Paper year* is the year the paper is published: the average is 1997 (minimum 1988 and maximum 2005). We then develop a series of gene paper variables based on a paper's authors and their affiliations. *Number of authors* counts the number of authors on the paper (mean = 7.28). *Number of addresses* is a count of the number of addresses that appear on the paper, measuring the number of different organizations involved in knowledge production (mean = 2.72). *U.S. address* is a binary variable capturing whether at least one of address is in the United States and reflecting the affiliation of the paper authors. *Public address* is a binary variable describing the condition under which at least one listed address is from a public institution (defined as universities, government research organizations and laboratories). We then construct the binary variable *private address* capturing whether there is at least one private sector address on the paper, coding pharmaceutical, biotechnology and other private sector corporations. *Impact factor* is a proxy for the journal quality in which the gene paper is published (mean = 9.75). Constructed by ISI and published in their annual Journal Citation Reports, it varies between 1 and 33.5. It is defined as the number of current year citations divided by the source items published in that journal during the previous two years. While the impact factor is re-calculated annually, the rank ordering of journals shows little or no variation over time; we therefore use 2005, the last paper publication year in our sample. The gene papers in our sample come from a small number of high quality journals. As shown in Table 4, only ten journals account for more

than fifty percent of the gene papers. These include general journals such as Science and Natures as well as more specialist journals: Genomics, Nature Genetics, Cell etc.

Insert Table 4 about here

Patent variables. A series of variables capture the temporal effects of a patent grant. *Patent application year* is the year in which the patent application is made (mean = 1997) while *patent grant year* is the year in which the patent is granted (mean = 2000). *Patent grant lag* is the elapsed time between patent application and patent grant (mean = 3.30). We also define the variable *patent in force*, which is a dummy variable equal to one for all years after a patent is granted and zero prior to patent grant. The mean of *patent in force* is 0.58, suggesting that our citation-year observations are almost equally distributed between years when patents are in operation and those when patents are not. *Patent window* is another dummy variable which is coded one during the year in which the patent is granted and zero otherwise (mean = 0.1).

We then developed a series of variables intended to denote characteristics of the patentees. *Number of inventors* and *number of assignees* measure the number of inventors (mean = 2.6) and assignees (mean = 1.1) listed on the patent respectively. *Public assignee* is defined in the same way as *public address* whether at least one of patent assignees is from a public (academic or government) institution (mean = 0.58). In addition, *all public assignee* denotes instances when all patent assignees are from a public institution (mean = 0.54). Similarly, *private assignee* is a binary variable to denote cases when at least one patent assignee is a private corporation (mean = 0.46) while *all private assignee* denotes all patent assignees from private corporations (mean = 0.42). *U.S. assignee* is set to one when at least one patent assignee is U.S. based (mean = 0.79).

We constructed additional patent variables in an attempt to capture various characteristics of the IP rights granted in the patent. *Patent scope*, following Lerner (1994), is defined as the number of national classes in which the patent is categorized (mean = 6.20) while the *number of*

claims in the patent (mean=15.61) is one measure of patent strength (Scotchmer, 1991). To analyze the impact of increase in scope from the mean, we constructed the measure based on the actual number of classes minus the mean number of classes. To analyze the impact of increasing strength, our measure captures the actual number of claims minus the mean number of claims.[11]

Patent-gene variables. As described above, the genetics setting allows us to measure the patent thicket directly. For our sample of 1,279 gene patent-paper pairs, each paired patent grants rights over one or more genes (there are 2,637 genes in our sample) and is the initial patent covering the gene paper. From the total gene patent population of 4,270 patents, we count the total number of patents eventually associated with each gene in the focal pair. We then break our count into five categorical variables: *gene patent 1* is a binary variable which is equal to one for patented genes that are claimed by one patent including the focal patent (mean = 0.335), *gene patent 2 to 4* is for genes with 2 to 4 patents (mean = 0.523), *gene patent 5 to 7* for genes with 5 to 7 patents (mean = 0.099), *gene patent 8 to 10* for genes with 8 to 10 patents (mean = 0.025) and finally *gene patent 11 to 20* for genes with 11 to 20 patents (mean = 0.017).[12] Table 5 gives the distribution of genes by number of times patented.

Insert Table 5 about here

To capture the degree of fragmentation of the gene(s) claimed by a particular patent *i*, we develop and construct the *patent-gene fragmentation index*. We first calculate a measure of *gene fragmentation* (for each gene) based on its Herfindahl index (*herfgene*): the sum of the squares of the shares of gene of each individual ownership (private firm or public organization). As more

---

[11] We analyzed both the impact of increase in patent scope and the impact of increase in scope from the mean (or positive deviation). The regression results are similar across both procedures. Similarly, we analyzed both the impact of increases in patent strength and the impact of increases in strength from the mean (or positive deviation). The regression results are similar across both procedures. We report the latter in Table 7.

[12] As genes claimed by more than 10 patents (i.e. 11 to 20 patents) represent only about 1.7 percent of the total observations in our sample, we have aggregated them into one category.

than one gene is associated with many of our patent-paper pairs, we then develop the *patent-gene fragmentation index* for each patent:

$$\sum_{j=1}^{ngene_i}(herfgene_j)/ngene_i \qquad (1)$$

where *j* indexes the number of genes, *ngene,* claimed by the particular patent *i* and *herfgene* is the computed Herfindahl index for each gene claimed by the same patent. This weights gene fragmentation across all the genes in the gene patent-paper pair. There is *more* fragmentation as the index approaches zero. The mean of the patent-gene fragmentation index is 0.67 (min 0.11 - most fragmented and max 1 - least fragmented). In order to make it more intuitive, we subtract this index from unity to denote more fragmentation with increasing index value (i.e. a value of 0 denotes no fragmentation while 0.89 denotes most fragmentation). To analyze the impact of increases in patent-gene fragmentation from the mean, we construct an additional measure by subtracting (1-mean fragmentation index) from (1-fragmentation index). The resulting measure is simply mean fragmentation index minus fragmentation index.[13]

We also explored the types of genes in our sample, specifically whether there is a link between the focal gene and a specific human disease at the time of its patent grant, as a proxy for its immediate usefulness and potential for commercialization. We did this by determining whether at least one of the genes in the gene patent-paper pair is listed in OMIM as expressing a human disease related phenotype – *OMIM gene* – the mean of 0.25 suggests about 25% of genes in our sample are linked to a human disease. Likewise *cancer gene* captures whether the focal gene is listed in the Sanger Institute Cancer Gene Census as a known cancer gene (mean = 0.076). Finally *disease gene* is a binary variable with captures whether the focal gene is listed

---

[13] Again, we analyzed both the impact of increases in fragmentation and the impact of increase in fragmentation from the mean fragmentation (or positive deviation). The regression results are similar across both procedures. We report the latter in Table 8.

either in OMIM or in Cancer (mean=0.27). Table 6 shows the linkage to disease and other key

patent characteristics for the top twenty most patented genes.

```
-------------------------------
```
Insert Table 6 about here
```
-------------------------------
```

Model Specification and Estimation

Our dependent variable is the annual count of citations received by each gene paper in peer-

reviewed journals (tracked by ISI) in every year since its publication. As this is a highly right-

skewed count variable that takes on non-negative integer values, we use a nonlinear regression

approach to avoid heteroskedastic, non-normal residuals (Hausman, Hall, & Griliches, 1984).

There are two ways to deal with the discrete nature of such count data: the Poisson regression

model (PRM) or the negative binomial regression model (NBRM), a generalized form of the

Poisson regression (Hausman, Hall, & Griliches, 1984). The Poisson assumes the conditional

mean of the outcome is equal to the conditional variance. However, as our dependent variable

exhibits over-dispersion with conditional variance significantly greater than the conditional mean

(Cameron & Trivedi, 1998),[14] this assumption is violated.[15] Hence, we employ the negative

binomial regression model, which overcomes the problem of over-dispersion by assuming a

gamma distribution for the conditional mean of the dependent count variable, and therefore

allows the conditional mean and variance to vary.[16]

We also assumed that the annual forward citations will depend on the specific calendar year

in which citations are being accrued (*citation year*) and the number of years since the gene paper

was published (*paper age*). As we are interested in whether the grant of the paired patent

---

[14] In the likelihood-ratio test, H_1: $E(y_{it}) < Var(y_{it})$ is supported.

[15] Note also that the standard errors from the Poisson regression model may be biased downwards resulting in spuriously large z-values (Cameron & Trivedi, 1986). The z-tests may over-estimate the significance of the variables in the case of over-dispersion in the data (Long, 1997). The results of the Hausman (1978) test also support the use of fixed effects negative binomial regression model.

[16] In our data, the goodness-of-fit test rejected the Poisson distribution assumption and indicated a zero-inflated distribution, showing further support for the negative binomial regression model.

changes the annual citation count, we include our main explanatory variable, *patent in force*, in the marginal effects equation (2). We also include the variable *patent window* allowing us to account for the fact that in the actual grant year of the patent, the impact of IP rights may be noisy. In equation (2), we rely on the observable characteristics of the papers to capture paper-by-paper differences in the underlying annual citation trend:

$$FC_{i,t} = f(\varepsilon_{i,t}; \alpha \text{patent\_window}_{i,t} + \beta \text{patent\_in\_force}_{i,t} +$$
$$\varphi \text{number\_of\_authors}_i + \xi \text{number\_of\_addresses}_i + \delta \text{US\_address}_i + \eta \text{public\_address}_i$$
$$+ \mu \text{impact\_factor}_i + \chi \text{paper\_age fixed effects}_{t\text{-paper\_year}} + \psi \text{citation\_year fixed effects}_t) \quad (2)$$

From equation (2), it is possible to develop a further specification. Instead of observable paper characteristics, equation (3) uses paper fixed effects for each paper to account for variation across individual papers.

$$FC_{i,t} = f(\varepsilon_{i,t}; \alpha \text{patent\_window}_{i,t} + \beta \text{patent\_in\_force}_{i,t} + \lambda \text{paper fixed effects}_i +$$
$$\chi \text{paper\_age fixed effects}_{t\text{-paper\_year}} + \psi \text{citation\_year fixed effects}_t) \quad (3)$$

In both equations we can test whether the citation rate to a paper changes after the paired patent is granted, accounting for fixed differences in the citation rate across papers with different observable characteristics (2) or across each paper (3). Using these two models for the assessment of the impact of patent grant, we then explore how patent strategy, including the various characteristics of the patent right, affects the supply of public scientific knowledge. In each case, we examine the interaction effects between *patent in force* and the particular patent or knowledge characteristics of interest represented by the last term in equation (4).

$$FC_{i,t} = f(\varepsilon_{i,t}; \alpha \text{patent\_window}_{i,t} + \beta \text{patent\_in\_force}_{i,t} + \lambda \text{paper fixed effects}_i + \chi \text{paper\_age}$$
$$\text{fixed effects}_{t\text{-paper\_year}} + \psi \text{citation\_year fixed effects}_t + \gamma \text{patent\_in\_force interactions}_{i,t}) \quad (4)$$

RESULTS

Our first analysis focuses on the impact of patent grant on the annual citation rate of paired papers. In all our Tables, we report the coefficients as incidence rate ratios (IRR) which can be

derived by exponentiating the coefficients, β_k of the independent variable x_k of the negative binomial regression models. In our case, the IRR can be interpreted as the factor change in annual citations received in a given year due to a unit increase in the regressor. For example, an IRR of 1.03 in the coefficient indicates a 3% increase in the dependent variable for a unit increase in the independent variable, all else equal.

Table 7 shows an estimate of a series of models using the negative binomial specifications as outlined above. Model 7-1 is a baseline model that estimates the annual citation count including fixed effects for *paper age* and *citation year* to account for the time trend in the citations and then including a series of observables on the gene papers including *number of authors, number of addresses, U.S. address* and *public address*. We also account for the quality of the journal in which gene papers are published, using the journal *impact factor*. In Model 7-2, we include the same variables and add in the post-patent grant effects using *patent window* and *patent in force*. In this model, which is a test of Hypothesis 1, we find that the grant of a gene patent negatively impacts the rate of follow-on publications as shown by the 17% decline in expected rate of forward citations to the paired gene paper (significant at the 0.1% level). This result is supported (although the effect is smaller) in the more stringent test provided in Model 7-3 which uses the full differences-in-differences estimate with a separate *paper* fixed effect for every article and the complete set of *paper age* and *citation year* fixed effects. In this rigorous specification we find the estimated negative impact of IP rights on forward citations to the paired paper is 5% (at the 1% level). This estimate is identified through the within-article comparison between the pre- and post- gene patent grant citation levels.

Insert Table 7 about here

Impact of patent scope and strength. To shed light on this effect, we analyze the degree to which patent characteristics contour the impact of patent grant by interacting the *patent in force* dummy with such characteristic measures as the *patent scope* and *patent strength*. In Model 7-4, we find that there is a statistically significant, although quite modest, impact of broader (scope) gene patents on subsequent citations to the gene paper (both relative to the mean and in absolute terms), with an incremental decline of about 1% (at the 0.1% level) on subsequent citations for every unit increase in number of patent class. Hypothesis 2 is therefore supported but with only a modest quantitative effect. Model 7-5 repeated this analysis for the strength of patent in terms of the number of legal claims. We find that there is no significant effect on subsequent citations of the publication as the patent is increasing in its number of claims (both relative to the mean and in absolute term). Therefore, hypothesis 3 is not supported suggesting that the number of claims in a given patent do not have an impact on how scientists engage in follow-on research on patented papers. Finally, Model 7-6 incorporates *patent in force,* the interaction effects of *patent scope* and *patent strength*. Results are stable and consistent with Models 7-3 to 7-5. The importance of patent scope in shaping the impact of patent enforcement on public knowledge production not only speaks to the responsiveness of follow-on researchers to differences in patent scope but most likely to differences in enforcement.

Impact of patent ownership. Another feature of the gene patenting debate that relates to enforcement highlights the relative role of public versus private patent holders in stifling follow-on public knowledge production via aggressive enforcement (or enforcement threats). On the one hand, the debate over patenting argues that academic patent owners are particularly problematic and largely contributes to the anti-commons effect (Heller & Eisenberg, 1998). On the other hand, much of the attention rests on private sector patentees in the gene patent debate. Our next set of analyses explores these two different possibilities and attempt to adjudicate between then by testing hypothesis 4. We estimate how the characteristics of the patent assignees

(and its attendant property rights) influence the rate of forward citation through a set of interaction effects between the variable *patent in force* and dummy variables for the patent assignee type as shown in Table 8.

Insert Table 8 about here

Model 8-1 compares the interaction between *public assignee* and *patent in force* with that of *no public assignee* and *patent in force*. Model 8-2 provides comparison of the interaction between *private assignee* and *patent in force* with that of *no private assignee* and *patent in force*. Model 8-3 checks on models 8-1 and 8-2 by examining the interaction effect between *all public assignee* and *patent in force* with that of *all private assignee* and *patent in force*. Model 8-1 shows a significant (at the 0.1% level) and negative impact of about 8% for *no public assignee* interaction, while the *public assignee* interaction is not significant. Model 8-2 shows a significant (at the 0.1% level) and negative impact of about 9% on *private assignee* interaction while *no private assignee* interaction is not significant. Model 8-3 forces the omitted category to be mixed and shows a significant (at the 1% level) and negative effect of 6% on *all private assignee* interaction while *all public assignee* interaction is not significant. Taken together, these findings strongly support hypothesis 4 and suggest that for the production of public knowledge (in human genetics), the main impact of patenting arises through private sector gene patents.[17]

Impact of complexity of patent landscape. Models 8-4 and 8-5 examine the role of the patent landscape in influencing the expected citation rates of gene papers, with a particular emphasis on the characteristics of ownership fragmentation. First, however, we examine the

---

[17] To check and insulate our results against any possibility that the interaction effects in a non-linear model are not the same as their cross-partial derivatives, we performed additional regressions similar to the one described in Model 7-3 on split samples separately for Models 8-1 to 8-8 (except Model 8-5). For example, in Model 8-1, we performed the regression in the sub-sample with *public assignee* only (7718 observations) and then another regression on the sub-sample with *no public assignee* only (5112 observations). We repeated this procedure for the remaining models. These split sample regressions yield consistent and robust results as those shown in Table 8 and our findings are unchanged across the models.

interaction of patent grant with characteristics of the patent thicket regardless of the ownership arrangements. Recall that our interaction term separates gene papers into different groups according to the number of associated gene patents that claim the underlying gene. Model 8-4 shows that patent grant over genes associated with 1 patent significantly depresses the citations by more than 7% while genes patented 5 to 7 times show a significant post-grant decline of more than 9% – a modest 2% increase due to denser "patent thickets" compared to genes with only one patent. However, impacts of genes patented 2 to 4 times, 8 to 10 times and 11 to 20 times are not statistically significant. While the 2% difference provides weak support for Hypothesis 5, the underlying relationship between gene patent grant and long-run public knowledge production is clearly not linear in the number of gene patents and does not increase smoothly with denser patent thicket. We therefore turn to an analysis of the ownership fragmentation of such patent thickets to examine whether ownership complexity contours the impact of patents on long-run knowledge production. When we analyze the impact of increase in fragmentation in Model 8-5, we find strong support for hypothesis 6. Specifically, we find that over and above the baseline decline in expected citations of 5% (Model 7-3), there is an incremental 7% significant decline (at the 5% level) in follow-on knowledge production for every unit increase in fragmentation of the patent thicket (relative to the mean and in absolute terms)[18]. For example, for any given patent, an increase in ownership from one to two organizations for a claimed gene would result in an incremental 3.5% decline in the forward citations of the paired paper.

Impact of usefulness of patented knowledge. In our final set of models, we examine the ways in which the characteristics of the knowledge disclosed in the gene papers and their paired patents have an impact on the degree to which patent grant affects the expected rate of forward citations. We do so by interacting the *patent in force* variable with a set of variables designed to

---

[18] In our analysis of the impact of increase in fragmentation using the measure presented in Equation 1, the regression result (available upon request) also showed a 7% significant decrease as fragmentation increases. Thus, our findings are consistent.

capture key genetic knowledge characteristics that proxy for the immediate usefulness and commercial importance of the gene. Model 8-6 compares the interaction between *OMIM gene* and *patent in force* with that of *no OMIM gene* and *patent in force*. There is about 8% significant decline (at the 0.1% level) for OMIM genes compared to no statistically significant decline in annual forward citations for non-OMIM genes. Model 8-7 does the same for *cancer gene* and *no cancer gene*. We find that there is a statistically significant impact of patent grant for both categories of gene, but the decrease is 11% for cancer genes versus 4% for non-cancer genes– a statistically significant difference of 7%. Finally, model 8-8 compares the interaction effects of *disease gene* versus *no disease gene*. We find a significant decline of 6% (at the 1% level) for disease gene compared with a decline of 4% for non-disease gene (at the 5% level). This difference of 2% is statistically significant. Taken together, the more immediately useful and relevant and commercializable the patented (genetic) knowledge, the more negative the impact of gene patent grant on subsequent published citations of that knowledge. Hypothesis 7 is therefore supported and we can clearly state that the negative impact of patent grant is centered on genes and genetic knowledge that is critical for human diseases.[19]

DISCUSSION

This paper provides the first large-scale systematic evidence of the impact of patenting on the long-run supply of public (published) knowledge in human genetics. Prior researchers who consider these questions typically take a narrow perspective: a single case study (e.g. Orsi and Coriat, 2005 on the BRCA1 diagnostic tests), a single sector (e.g. Cho, Illangasekare, Weaver, Leonard, & Merz, 2003 on genetic testing) or only litigated patents (Holman, 2007). Using the

---

[19] As an additional check against potential co-linearity among the fixed effects, we also performed the "fully interacted" specification on regression Models 7-3 to 7-6 and 8-1 to 8-8 i.e. instead of *paper* fixed effects, *paper age* fixed effects and *citation year* fixed effects in the models, we include *paper* fixed effects and the full set of *paper year-paper age* interaction dummies. The results are consistent and robust across all the models as those reported in Tables 7 and 8 – the coefficients have similar direction and almost identical magnitude (available upon request).

differences-in-differences approach, we confirm prior results arguing for the negative impact of patent grant on the future production of the public knowledge stream.

More importantly, we significantly extend our understanding of these issues by articulating how patent strategies – patent scope, patent ownership, complexities of the patent landscape and commercial relevance of patented private knowledge – have a causal negative impact on the long-run production of public knowledge. Taken together, our results suggest that patent strategies pursued by firms and others, in their attempt to build private knowledge streams, are tightly coupled to the long-run public knowledge stream. While prior scholarship articulates a positive relationship between these streams, our work argues that by taking a reciprocal approach and considering the impact of private knowledge production strategies on the long-run production of public knowledge, we uncover some of the unintended consequences of firms' patent strategies.

Our study provides policymakers and managers with novel evidence that can be brought to bear in the ongoing gene patent debate. Our analysis confirms that the public and private knowledge streams are intertwined, with patent-paper pairs as a prevalent disclosure strategy pursued by firms and by academics in human genetics that allows them to incorporate their genetic knowledge into both the public and private institutional spheres. It also shows that gene patent-paper pairs are not exclusively a disclosure strategy in the domain of either academic or for-profit organizations. Instead, both types of organizations contribute (at least some of) their genetic knowledge in the form of patent-paper pairs. Finally, pairs are not only a form of disclosure for obscure, poorly understood genes, but also for disease genes (about 27% of our sample), including those implicated in cancer. Consequently, pairs are an essential tool shaping the production, disclosure and accumulation of genetic knowledge in our economy.

Unlike previous studies of gene patenting, our analysis is comprehensive: the gene patent-paper pairs we analyzed disclose 2,637 gene sequences – more than 11% of the known human

genes – ranging from controversial to less well-known genes. From this broad perspective, we estimate that the negative impact of patent grant on the future public knowledge production is about 5% (in the most stringent differences-in-differences specification). This result is consistent in direction and magnitude with prior research using patent-paper pairs (Murray & Stern, 2007a), but more robust because it provides evidence over a much longer time-horizon, using a larger and more diverse sample of publications. Our empirical approach also makes a tighter link between the patent and the paper by focusing on pairs claiming property rights over gene sequences.

A strict interpretation of our results suggests follow-on genetic researchers forego about one in ten research projects (or more precisely research publications) through the causal negative impact of the gene patent grant. Of course, as noted earlier, it is possible that researchers are continuing their research and contributing to the public knowledge stream while engaging in strategic citing behavior (citing papers not covered by the gene patent). However, by showing the extent to which the patent grant effect is contingent upon patent scope, patent ownership and the fragmentation in ownership, we argue that these outcomes are driven by patent enforcement issues and complexities in bargaining and transactions costs rather than strategic citing. It is hard to envision strategic citing behavior being so responsive to the details of the patent landscape, particularly given the evidence that researchers do not directly analyze the relationship between patents and their research inputs (Walsh et al., 2005). However, the decline does not necessarily mean knowledge production has halted. Instead, follow-on researchers may be increasing secrecy – which would show up as fewer forward citations to the paired gene paper.

Four further elements of our results elaborate and deepen our understanding of the role of patents on the public stream of genetic knowledge. First, the effect of gene patents is centered on patents owned by the private sector. In contrast to the prior literature, which finds that public sector patents have a greater negative impact on forward citations, we found patents assigned to

private sector firms have a more negative impact – a decline of 6% to 9% compared with only 0% to 3% for academic patents. We interpret the greater impact to mean that private sector enforcement (or anticipated to enforcement) on follow-on researchers was more aggressive or is anticipated to be more aggressive by these firms during the period of our analysis which spans 1988 to 2006. If knowledge accumulation (particularly by firms) is most effectively undertaken when follow-on researchers contribute to the public knowledge stream, we interpret the 8% decline in forward citations that accompanies private sector gene patenting as a net loss to long-run public knowledge production, a condition of particular concern to firms themselves. This finding contradicts previous scholarship using a sample of biotechnology patent-paper pairs which shows that over the shorter period (1997 to 2003) paired public sector patents have a more negative impact on forward citations than private sector patents (Murray & Stern, 2007a). In recent follow-on work, these authors show that the public sector effect is ameliorated over time (Murray & Stern, 2008), consistent with our finding that publicly-owned patents have little long-run impact. Furthermore, in gene patenting there is considerable supporting evidence that the private sector has been aggressive in its enforcement strategies with other firms and with public sector organizations, particularly academic medical centers. Indeed, in additional analysis we find a particularly substantial decline in medical-center citations.

Second, the notion that the negative impact of gene patents is due to patent enforcement is supported by our key result – the impact of patents on long-run public knowledge is increasing in the scope of patents. This evidence shows that it is strong enforcement of broader patents or the probability of strong patent enforcement that drives the dampening effect of patent grant on public knowledge production.

The third element of our results highlights the ways in which the patent landscape contours the long-run production of genetic knowledge. Beyond single gene patents, scholars have debated whether and to what extent patent thickets and patent fragmentation impede follow-on

innovation. While Ziedonis (2004) finds evidence that complex patent thickets limit future patent contributions to the private knowledge stream (in the semiconductor industry), no empirical evidence speaks directly to the relationship between the patent landscape and public knowledge production. We take advantage of the specificity of gene sequences to develop the first large-scale empirical evidence of this type. We show that while the relationship between patents thicket size and knowledge production is noisy (the results are directionally as we predict), increasing ownership fragmentation contributes significantly to the negative impact of patent grant. This suggests that while ownership concentration is an important source of competitive advantage for owners of the patent portfolio, fragmentation is more problematic for follow-on contributors to public knowledge presumably because of the complexities and costs of navigating and negotiating with many patent assignees in a fragmented patent thicket.

Lastly, beyond characteristics of individual patents and patent landscape produced around them, we also find that the negative effect of patents on follow-on public knowledge production is greatest for genes closely linked to human disease i.e. that are more immediately useful and with greater commercial potential. While we would expect that gene papers on disease-related genes are more highly cited, our methodology allows us to distinguish between this "levels" effect and the effect of gene patent grant on annual citations. When combined with the fact that such patents are more likely to be the subject of aggressive enforcement tactics such as "cease and desist" letters or litigation threats, this is perhaps our most important finding because it suggests the negative effect of patents lies right at the heart of the fight to improve human health.

Implications

Our findings have a number of implications for policy and management (we turn to the broader implications for scholars in our final subsection). Policymakers should take our analysis as evidence that privately owned gene patent rights might have a detrimental long-run impact on

specific lines of public genetic knowledge. To the extent that our result on genetics are robust in other knowledge settings, this shifts the policy debate away from the role of academic patenting and instead highlights the relationship between industry and all those engaged in public knowledge production. It suggests that policymakers continue to explore the appropriate scope and strength of gene patents and provide greater clarity over their enforcement. There are also policy choices that can ameliorate the impact of patents on researchers in academia. Specifically, a federally mandated experimental use exemption may be an important mechanism to facilitate follow-on knowledge production (Dreyfuss, 2004). To date, industry has been opposed to the idea that academic researchers be formally exempt from patent rights (notwithstanding the fact that many academics take this right *de facto*). The potential for a firm's patent rights to stifle long-run public knowledge production may provide firms with more convincing evidence they need to support an experimental exemption and therefore potentially lift the stifling impact of their patent rights.

For managers, our results suggest that there is a more complex and bi-directional link between a firm's private knowledge stream and the public stream. While public knowledge certainly makes a critical contribution to a firm's private knowledge stream (indeed on average, the industry-assigned patents in our sample cite 434 non-patent references to public genetic knowledge), a firm's patents, and the patent landscape that emerges, contour and stifle the public knowledge stream over the long-run. This is consistent with a recent quantitative study illustrating the positive impact on public knowledge production – over 30% increase in publication citation when a firm makes its patented research materials more open and widely accessible to academic researchers (Murray, Aghion, Dewatripont, Kolev, & Stern, 2008). As noted above, this stifling effect may be effectively mitigated by an experimental use exemption. Furthermore, firms may be able to ameliorate these effects by making changes in licensing requirements, practices and policies. Patent pools or cross-licensing arrangements associated

with particularly "useful" knowledge may reduce the negative effect of the complex patent landscape. We also suggest that access to *information* defining patent rights (held by public or private entities) on a gene-by-gene basis may help improve the impact of patent thickets. At present there is no systematic, accessible and regularly updated source of information that allows a public or private sector scientists to establish whether and what patent rights pertain to specific genes. The provision of such a service would be an important starting point for navigating a complex patent landscape and reducing the uncertainty with already probabilistic patents.

As an important caveat, we should note that our analysis does not provide evidence for the potentially beneficial effects of patent rights as they are traditionally justified, for example as an incentive for commercial investments (Mazzoleni & Nelson, 1998), in the market for ideas (Gans et al., 2008) etc. In the long-run, the tradeoff between patents' negative effects on follow-on innovators and their positive incentive effect must be adjudicated.

Future Research Agenda

Our study recognizes the richness of the public and private institutional spheres in which knowledge is embedded (Dasgupta & David, 1994). Beyond the choice of whether to publish ideas and to file for patents, the flexibility of these institutional arrangements provides individuals, firms, and communities with a wealth of opportunities to construct new approaches to knowledge production, disclosure and accumulation (Murray & O'Mahony, 2007). This suggests a rich new research agenda placing knowledge work into its institutional context while at the same time highlighting the important, and sometimes overlooked, role played by firm strategy, particularly concerning IP decisions. More specifically, it guides research questions for management scholars at the intersection of three literatures: strategic management of intellectual property; institutional theory, particularly focused on the sociology of law and norms; and organization theory pertaining to knowledge work. We examine each of these in turn.

The patenting and publishing behaviors described in human genetics bring two institutional spheres together and, at times, into collision and conflict. This collision is not unique to the life sciences. Instead, it animates knowledge-based firms and knowledge workers in music, software, and beyond and holds important long-run implications for public and private knowledge production (Heller, 2008; Lessig, 2004; Scotchmer, 2004). The richness of these interactions has been the focus of recent institutional scholarship, highlighting the ways in which these two institutional spheres shape knowledge production in academia (Biagioli, 2002; Colyvas & Powell, 2006; Kleinman, 2003; Murray, 2008; Owen-Smith & Powell, 2003). However, these studies may have devoted too little attention to the complexity and flexibility of the private property sphere. All too often, this literature conceptualizes the legal environment as rigid and inflexible, particularly in comparison to the informal norms of the public commons. This is contrary to the insights from the sociology of law, which, while not focused on IP law, argues that individuals, communities and particularly firms are "immersed in a sea of law" (Dobbin & Kelly, 2007; Edelman, 1992; Kelly & Dobbin, 1999) and have many opportunities to influence their legal context. This perspective suggests that just as scholars now examine how contributors to the public commons constitute informal norms (Fauchart & von Hippel, 2008; Oliar & Sprigman, 2008), we should also examine how firms (and other actors) construct the norms and practices of the private property sphere using the rich theoretic lenses of the law and society tradition.

For scholars interested in the causes and consequences of firms' IP decisions, our paper argues that while it is critical to understand the ways in which IP strategies provide for property rights and competitive advantage (Gans & Stern, 2000; Somaya, 2003; Teece, 1986), we must also attend to their impact on long-run knowledge production (Scotchmer, 2004; Ziedonis, 2004). Specifically, IP scholars might consider private property and public commons institutions. This focus resonates with the current interest in "open innovation" strategies

(Chesbrough, 2003), but provides a broader, more institutionally focused framework for considering such issues.

A central question raised by taking this approach is to understand whether and how firms' strategies for engagement with the public commons i.e. using published literature, providing open source software code, or limiting IP enforcement, shape the long-term production of a wide range of public goods. It is not enough for firms to assume that the public commons are robust to their intervention. Specifically, our work suggests that more than simply a public good to be appropriated into the private knowledge stream (Cohen & Levinthal, 1990), the public commons are highly sensitive to patent strategy and the patent landscape that is produced. Such an agenda turns the traditional approach to public goods on its head; current analysis on open innovation generally examines how firms capture the value of distributed innovation generated by individuals and communities who freely reveal their knowledge (Harhoff, Henkel, & von Hippel, 2003; Lakhani & von Hippel, 2003). Before we conclude that such strategies are a clever way of getting something for nothing, we must consider whether they limit the long-run production of the very public goods upon which many firms have come to rely. Current studies of for-profit firms' attempts to appropriate public goods hint at the complexity of the link between public and private knowledge production. For example, some firms seeking advantage from open-source contributions to software code find that their strategies decrease the rate of code contributions (von Hippel & von Krogh, 2006).

What lies at the heart of these transformations? The last element of our proposed agenda calls for a deeper understanding of knowledge work, of how it is embedded in the public and private institutional spheres, and how it is influenced by firm strategies as they operate within these spheres. Why and how, for example, do knowledge workers contributing to the public commons respond to firms' attempts to contribute to or appropriate their work? Do these responses change the types of individuals making such contributions or the types of

contributions? Much of the existing scholarship on knowledge work assumes that the institutional and firm environments are relatively stable (Bechky 2003; Hargadon & Bechky, 2006; Hargadon & Sutton, 1997). However, the dynamic collisions between the public and private spheres suggest otherwise. This raises new questions that can only be explored when organizational theorists attend to both firm strategy and the complex institutional environment. Recent studies show that novel arrangements emerge as individuals, firms and other actors develop solutions that make it possible to engage in the daily practice of knowledge while living with the law and with other institutional opportunities and constraints (Meyer & Argyres, 2004). Some of these solutions seem to derive from complex and nuanced contracts that allow for mutual co-existence of knowledge communities with distinctive aims and goals (Murray & Stern, 2008). Others seem to be more organizational in nature, providing governance of the tensions through organizational devices such as committees etc. (O'Mahony & Bechky, 2008). However, there is still much to learn and the agenda will require scholarship bridging institutional theory, IP strategy and organizational theories of knowledge work.

REFERENCES

Abate, T. 2000. *Call it the gene rush – patent stakes run high*. San Francisco Chronicle, April 20, pp. A8.

Aghion, P., Dewatripont, M., & Stein, J. 2005. *Academic freedom, private-sector focus and the process of innovation.* Working paper no. 11542, National Bureau of Economic Research, Cambridge, MA.

Almeida, P., Dokko, G., & Rosenkopf, L. 2003. Startup size and the mechanisms of external learning: Increasing opportunity and decreasing ability? *Research Policy,* 32:301-315.

Andrews, E. L. 1991. *U.S. seeks patent on genetic codes, setting off furor.* New York Times, 10/21/1991.

Bechky, B. 2003. Sharing meaning across occupational communities: The transformation of knowledge on a production floor. *Organization Science,* 14: 312-330.

Biagioli, M. 2000. Replication or monopoly? The economies of invention and discovery in Galileo's observations of 1610. *Science in Context,* 13(3-4): 547-590.

Blumenthal, D. 1997. Withholding research results in academic life science: Evidence from a national survey of faculty. *Journal of the American Medical Association,* 277: 1224-8.

Borson, D. B. 1995. The human genome projects: Patenting human genes and biotechnology. Is the human genome patentable? *IDEA: The Journal of Law and Technology,* 35(4): 461-496.

Bush, V. 1945. *Science: The endless frontier.* Washington, D.C.: United States Government Printing Office.

Brown, J. S., & Duguid, P. 2001. Knowledge and organization: A social-practice perspective. *Organization Science,* 12(2): 198-213.

Cameron, C. A., & Trivedi, P. K. 1986. Econometric models based on count data: Comparisons and applications of some estimators and tests. *Journal of Applied Econometrics,* 1: 29-54.

Cameron, C. A., & Trivedi, P. K. 1998. *Regression analysis of count data.* New York, NY: Cambridge University Press.

Campbell, E. G., Clarridge, B. R., Gokhale, M., Birenbaum, L., Hilgartner, S., Holtzman, N.A., & Blumenthal, D. 2002. Data withholding in academic genetics: Evidence from a national survey. *Journal of the American Medical Association,* 287(4): 473 - 479.

Chesbrough, H. 2003. *Open Innovation.* Boston, MA: Harvard Business School Press.

Cho, M. K., Illangasekare, S., Weaver, M. A., Leonard, D. G. B., & Merz, J. F. 2003. Effects of patents and licenses on the provision of clinical genetic testing services. *Journal of Molecular Diagnostics,* 5(1): 3-8.

Cockburn, I., & Henderson R. 1998. Absorptive capacity, coauthoring behavior, and the organization of research in drug discovery. *Journal of Industrial Economics,* 46(2): 157-182.

Cohen, W., & Levinthal, D. 1990. Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly,* 35(1): 128-152.

Colyvas, J. A., & Powell, W. W. 2006. Roads to institutionalization: The remaking of boundaries between public and private Science, *Research in Organizational Behavior,* 21:305-53.

Crichton, M. 2006. *Next.* New York, NY: Harper Collins.

Dasgupta, P., & David, P. A. 1994. Towards a new economics of science. *Research Policy,* 23: 487-521.

de Solla Price, D. J. 1965. Networks of scientific papers. *Science,* 149(3683): 510-515.

Dobbin F., & Kelly, E. 2007. How to stop harassment: The professional construction of legal compliance in organizations. *American Journal of Sociology,* 112(4): 1203-1243.

Dreyfuss, R. 2004. Protecting the public domain of science: Has the time for an experimental use defense arrived? *Arizona Law Review,* 46: 457-472.

Ducor, P. 2000. Intellectual property: Coauthorship and coinventorship. *Science,* 289: 873-875.

Duke Law and Technology Review. 2001. The fate of gene patents under the new utility guidelines. February 28: 008. Accessed from http://www.law.duke.edu/journals/dltr/ARTICLES/2001dltr0008.html

Eisenberg, R. 1996. Public research and private development: Patents and technology transfer in government-sponsored research. *Virginia Law Review,* 82(8):1663-1727.

Fauchart, E., & von Hippel, E. 2008. Norms-Based intellectual property systems: The case of French chefs. *Organization Science,* 19(2): 187-201.

Ferber, M. A. 1988. Citations and networking. *Gender and Society,* 2(1): 82-89.

Fleming, L., & Sorenson, O. 2004. Science as a map in technological search. *Strategic Management Journal,* 25: 909-928.

Furman, J., & Stern, S. 2006. *Climbing atop the shoulders of giants: The impact of institutions on cumulative research.* Working paper, National Bureau of Economic Research, Cambridge, MA.

Gans, J., Hsu, D., & Stern, S. 2008. The impact of uncertain intellectual property rights on the market for ideas: Evidence from patent grant delays. *Management Science,* 54(5): 982-997.

Gans, J., Murray, F., & Stern, S. 2008. *Patents, papers and secrecy: Scientific disclosure as a negotiation between scientists and those who fund them.* Working paper, Massachusetts Institute of Technology Sloan School of Management, Cambridge, MA.

Gans, J., & Stern, S. 2000. Incumbency and R&D incentives: Licensing the gale of creative destruction. *Journal of Economics and Management Strategy,* 8(4): 484-511.

Gans, J., & Stern, S. 2003. The product market and the market for 'ideas': Commercialization strategies for technology entrepreneurs. *Research Policy,* 32(2): 333-350.

Garfield, E. 1955. Citation indexes for science: A new dimension in documentation through association of ideas. *Science,* 122(3159): 108-111.

Gittelman, M., & Kogut, B. 2003. Does good science lead to valuable knowledge? Biotechnology firms and the evolutionary logic of citation patterns. *Management Science,* 49(4): 366-382.

Greenfield, D. L. 2006. Greenberg v. Miami Children's Hospital: Unjust enrichment and the patenting of human genetic material. *Annals of Health Law,* 15(2): 213-249.

Hall, B. H., Jaffe, A. B., & Trajtenberg, M. 2001. *The NBER patent citation data file: Lessons, insights and methodological tools.* Working paper no. 8498, National Bureau of Economic Research, Cambridge, MA. Available at SSRN: http://ssrn.com/abstract=285618.

Hansen, M. T. 1999. The search-transfer problem: The role of weak ties in sharing knowledge across organization subunits. *Administrative Science Quarterly,* 44: 82-111.

Hagström, W. O. 1965. *The scientific community.* New York: Basic Books.

Hargadon, A. B., & Bechky, B. A. 2006. When collections of creatives become creative collectives: A field study of problem solving at work. *Organization Science,* 17: 484-500.

Hargadon, A., & Sutton, R. I. 1997. Technology brokering and innovation in a product development firm. *Administrative Science Quarterly,* 42: 716-749.

Harhoff, D., Henkel, J., & von Hippel, E. 2003. Profiting from voluntary information spillovers: How users benefit from freely revealing their innovations. *Research Policy,* 32(10): 1753-1769.

Harhoff, D., & Reitzig, M. 2004. Determinants of opposition against EPO patent grants: The case of biotechnology and pharmaceuticals. *International Journal of Industrial Organization,* 22(4): 443-480.

Hausman, J. A. 1978. Specification tests in econometrics. *Econometrica,* 46: 1251-1271.

Hausman, J. A., Hall, B. H., & Griliches, Z. 1984. Econometric models for count data with an application to the patents-R&D relationship. *Econometrica,* 52: 909-938.

Heller, M. A. 2008. *Gridlock economy: How too much ownership wrecks markets, stops innovation and costs lives.* Philadelphia, PA: Basic Books.

Heller, M. A., & Eisenberg, R. S. 1998. Can patents deter innovation? The anti-commons in biomedical research. *Science,* 280 (5364): 698-701.

Henry, R. M., Cho, M. K., Weaver, M. A., & Merz, J. F. 2002. DNA patenting and licensing. *Science,* 297 (5585): 1279.

Henderson, R., & Cockburn, I. 1994. Measuring competence? Exploring firm effects in pharmaceutical research. *Strategic Management Journal,* 15: 63-84.

Hitt, M. 2005. Management theory and research: Potential contributions to public policy and public organizations. *Academy of Management Journal,* 48(6): 963-966.

Hoetker, G. & Agarwal, R. 2007. Death hurts but isn't fatal: The post-exit diffusion of knowledge created by innovative companies. *Academy of Management Journal,* 50(2): 446-467.

Holman, C. M. 2007. *The impact of human gene patents on innovation and access: A survey of human gene patent litigation.* Working Paper IPSC 2007. http://www.law.depaul.edu/institutes_centers/ciplit/ipsc/paper/Chris_HolmanPaper.pdf.

Holman, M. A., & Munzer, S. R. 2000. Intellectual property rights in genes and gene fragments: A registration solution for expressed sequence tags. *Iowa Law Review,* 85(3): 735-848.

Huang, K. G. 2006. *Innovation in the life sciences: The impact of intellectual property rights on scientific knowledge diffusion, accumulation and utilization.* Unpublished doctoral dissertation, Massachusetts Institute of Technology.

Huang, K. G., & Murray, F. 2008. *Do organizational choices shape scientific progress? The Human Genome Project as a policy experiment.* Working paper, Singapore Management University Lee Kong Chian School of Business, Singapore.

Jensen, K., & Murray, F. 2005. The intellectual property landscape of the human genome. *Science,* 310: 239-240.

Kelly, E., & Dobbin, F. 1999. Civil rights law at work: Sex discrimination and the rise of maternity leave policies. *American Journal of Sociology,* 105(2): 455-492.

Kleinman, D. L. 2003. **Impure cultures: University biology and the world of commerce.** Madison, WI: University of Wisconsin Press.

Klevorick, A., Levin, R., Nelson, R., & Winter, S., 1995. On the sources and significance of inter-industry differences in technological opportunities. *Research Policy,* 24(2): 185-205.

Knorr-Cetina, K. 1999. *Epistemic cultures: How sciences make knowledge.* Cambridge, MA: Harvard University Press.

Kogut, B., & Almeida, P. 1999. Localization of knowledge and the mobility of engineers in regional networks. *Management Science,* 45: 905-917.

Kogut, B., & Zander, U. 1992. Knowledge of the firm, combinative capabilities, and the replication of technology. *Organization Science,* 3(3): 383-397.

Kohler, R. E. 1994. Lords of the fly: "Drosophila" genetics and the experimental life. Chicago, IL: University of Chicago Press.

Krimsky, S. 2003. *Science in the private interest.* Lanham, MD: Rowman & Littlefield Publishing Group.

Lanjouw, J. O., & Schankerman, M. A. 2001. *Enforcing intellectual property rights.* Discussion paper no. 3093, Center for Economic and Policy Research, Washington, D.C.

Lakhani, K., & von Hippel, E. 2003. How open source software works: ``Free" user-to-user assistance. *Research Policy,* 32(6): 923-943

Latour, B., & Woolgar, S. 1979. *Laboratory Life: The Construction of Scientific Facts.* Beverly Hills, CA: Sage Publications.

Lemley, M. A., & Shapiro, C. 2005. Probabilistic patents. *Journal of Economic Perspectives,* 19(2): 75-98.

Lerner, J. 1994. The Importance of Patent Scope: An Empirical Analysis. *RAND Journal of Economics,* 25(2): 319-333.

Lessig, L. 2004. *Free culture: How big media uses technology and the law to lock down culture and control creativity.* New York, NY: The Penguin Press.

Lewis, T. R., & Yao, D. A. 1995. Some reflections on the antitrust treatment of intellectual property. *Antitrust Law Journal,* 63(2): 603-619.

Lim, K. 2000. *Basic research, applied research, and innovation in the semiconductor and pharmaceutical industries.* Unpublished doctoral dissertation, Massachusetts Institute of Technology.

Long, S. J. 1997. Count outcomes: Regression models for counts. In *Regression models for categorical and limited dependent variables:* 217-250. Thousand Oaks, CA: Sage.

MacKie-Mason, J. K. 2002. *What to do about unilateral refusals to license?* Paper submitted as testimony to FTC-DOJ Hearings on "Competition and Intellectual Property Law and Policy in the Knowledge-Based Economy," available at http://www.ftc.gov/opp/intellect/detailsandparticipants.htm#May%201

Merton, R. K. 1973. The normative structure of science. In *The sociology of science: Theoretical and empirical investigations:* 267-280 (Ch. 13). Chicago, IL: The University of Chicago Press.

Merz, J. F. 1999. Disease gene patents: Overcoming unethical constraints on clinical laboratory medicine. *Clinical Chemistry*, 45(3): 324-330.

Meyer, K. J., & Argyres, N. S. 2004. Learning to Contract: Evidence from the Personal Computer Industry, *Organization Science*, 15(4): 394-410.

Mowery, D., Nelson, R., Sampat, B., & Ziedonis, A. 2001. The growth of patenting and licensing by U.S. universities: An assessment of the effects of the Bayh-Dole act of 1980. *Research Policy*, 30: 99-119.

Mazzoleni, R., & Nelson, R. R. 1998. Economic theories about the benefits and costs of patents. *Journal of Economic Issues*, 32(4): 1031-1052.

Meyer, M. 2000. Does science push technology? Patents citing scientific literature. *Research Policy*, 29(3): 409-434.

Murray, F. 2002. Innovation as co-evolution of scientific and technological networks: Exploring tissue engineering. *Research Policy*, 31 (8-9): 1389-1403

Murray, F. 2008. *The oncomouse that roared: Hybrid exchange strategies as a source of productive tension at the boundary of overlapping institutions.* Working Paper, Massachusetts Institute of Technology Sloan School of Management, Cambridge, MA.

Murray, F., Aghion, P., Dewatripont, M., Kolev, J., & Stern, S. 2008. *Of mice and growth: The effects of openness on follow-on research.* Working Paper, Massachusetts Institute of Technology Sloan School of Management, Cambridge, MA.

Murray, F., & O'Mahony, S. 2007. Exploring the Foundations of Cumulative Innovation: Implications for Organization Science. *Organization Science*, 18(6): 1006-1021.

Murray, F., & Stern, S. 2007a. Do formal intellectual property rights hinder the free flow of scientific knowledge? An empirical test of the anti-commons hypothesis. *Journal of Economic Behavior and Organization*, 63(4): 648-687.

Murray, F., & Stern, S. 2007b. When ideas are not free: The impact of patents on scientific research. In A. B. Jaffe, J. Lerner & S. Stern, (Eds.), *Innovation Policy and the Economy*, 7:33-69. Cambridge, MA: MIT Press.

Murray, F., & Stern, S. 2008. *Learning to live with patents: Assessing the dynamic adaptation to the law by the scientific community.* Working Paper, Massachusetts Institute of Technology Sloan School of Management, Cambridge, MA.

Narin F, Hamilton, K, & Olivastro, D. 1997. The increasing linkage between U.S. technology and public science. *Research Policy*, 26(3): 317-330.

Nash, M. 2000. *Who owns the genome?* CNN, April, 10.

Nomaler, Ö., & Verspagen, B. 2007. Knowledge flows, patent citations and the impact of science on technology. Working paper 2007-022, UNU-MERIT.

O'Mahony, S., & Bechky, B. 2008. *The role of boundary organizations in managing the problem of incommensurability.* Working paper, University of California, Davis, Graduate School of Management, Davis, CA.

Oliar, D., & Sprigman, C. 2008. Intellectual property norms among stand-up comedians. *Virginia Law Review,* forthcoming.

Orsi, F., & Coriat, B. 2005. Are "strong patents" beneficial to innovative activities? Lessons from the genetic testing for breast cancer controversies. *Industrial and Corporate Change,* 14(6): 1205-1221.

Owen-Smith, J., & Powell, W. 2003. The expanding role of university patenting in the life sciences: Assessing the importance of experience and connectivity. *Research Policy,* 32(9): 1695-1711.

Posner, R. A. 2000. An economic analysis of the use of citations in the law. *American Law and Economics Review,* 2(2): 381-406.

Powell, W., Koput, K., & Smith-Doerr, L. 1996. Interorganizational collaboration and the locus of innovation: Networks of learning in biotechnology. *Administrative Science Quarterly,* 41(1): 116-145.

Resnik, D. 1998a. Conflicts of interest in science. *Perspectives on Science,* 6(4): 381-408.

Resnik, D. 1998b. Industry-sponsored research: Secrecy versus corporate responsibility. *Business and Society Review,* 99(1): 31-35.

Restaino, L. G., & Takeuchi, T. 2006. Gene patents and global competition issues. *Genetic Engineering and Biotechnology News,* 26(1).

Romer, P. M. 1994. The origins of endogenous growth. *Journal of Economic Perspectives,* 8(1): 3-22.

Rosenberg, N. 1974. Science, invention and economic growth, *Economic Journal,* 84(333): 90-108.

Rysman, M., & Simcoe, T. S. 2008. *Patents and the performance of voluntary standard setting organizations.* Forthcoming *Management Science* and Working Paper No. 05-22, NET Institute. Available at SSRN: http://ssrn.com/abstract=851245.

Sampat, B. N. 2005. *Do academic genomic patents curtail downstream research?* Working paper, Mailman School of Public Health, Columbia University, New York, NY.

Scotchmer, S. 1991. Standing on the shoulders of giants: Cumulative research and the patent law. *Journal of Economic Perspectives,* 5: 29-41.

Scotchmer, S. 2004. *Innovation and incentives.* Cambridge, MA: MIT Press.

Shapiro, C. 2001. Navigating the patent thicket: Cross licenses, patent pools, and standard-setting. In A. B. Jaffe, J. Lerner & S. Stern, (Eds.), *Innovation Policy and the Economy*, 1:119-50. Cambridge, MA: MIT Press.

Shockley, W. 1949. The theory of P-N junctions in semiconductors and P-N junction transistors. *Bell System Technical Journal*, 28: 435–89.

Somaya, D. 2003. Strategic determinants of decisions not to settle patent litigation. *Strategic Management Journal*, 24(1): 17-38.

Stern, S. 2004. Do scientists pay to be scientists? *Management Science*, 50(6): 835-853.

Stigler, G., & Friedland, C. 1975. The citation patterns of doctorates in Economics. *Journal of Political Economy*, Vol. 83, No. 3, pp. 477-507.

Teece, D. J. 1986. Profiting from technological innovation. *Research Policy*, 15(6): 285-305

The Independent. 2007. James Watson: Genetic disorder. October 20. Accessed from http://www.mindfully.org/GE/2007/James-Watson-Speaks20oct07.htm

Thursby J., & Thursby M. 2002. Who is selling the ivory tower? Sources of growth in university licensing. *Management Science*, 48(1): 90–104.

Tijssen, R. J. W. 2002. Science dependence of technologies: Evidence from inventions and their inventors. *Research Policy*, 31(4): 509-526.

Trajtenberg, M., Henderson, R., & Jaffe, A. B. 1997. University vs. corporate patents: A window on the basicness of invention. *Economics of Innovation and New Technology*, 5(1): 19-50.

Veggeberg, S. 1992. Controversy mounts over gene patenting policy. *The Scientist*, 6(9):1.

von Hippel, E., & von Krogh, G. 2006. Free revealing and the private-collective model for innovation incentives. *R&D Management*, 36(3): 295-306.

Walsh, J. P., Arora, A., & Cohen, W. M. 2003. Patenting and licensing of research tools and biomedical innovation. In S. Merrill, R. Levin & M. Meyers (Eds.), *Innovation in a Knowledge-Based Economy:* 285-340. Washington, D.C.: National Academies Press.

Walsh, J. P., Cho, C., & Cohen W. M. 2005. View from the bench: Patents and material transfers. *Science*, 309(5743): 2002-2003.

Weitzman, M. 1974. Free access vs. private ownership as alternative systems for managing common property. *Journal of Economic Theory*, 8: 225-234.

Ziedonis, R. H. 2004. Don't fence me in: Fragmented markets for technology and the patent acquisition strategies of firms. *Management Science*, 50(6): 804-820.

Zucker, L., Darby, M., & Brewer, M. 1998. Intellectual human capital and the birth of U.S. biotechnology enterprise. *American Economic Review*, 88(1): 290-306.

TABLE 1
Key Variable Definitions

Citation-Year Characteristics		
Name	**Definition**	**Source**
Annual cite	Number of citations made by later papers to the (focal) paper previously published in a given year	ISI
Total cite	Total number of citations accruing to a paper over its lifetime	ISI
Citation year	The year in which the forward citation is received	ISI
Paper age	Age of paper when a citation is made	ISI
Paper Characteristics		
Paper year	Year when paper is published	ISI
Number of authors	Number of authors appearing on the paper	ISI
Number of addresses	Number of unique addresses appearing on paper	ISI
U.S. address	Binary variable (1/0) denoting at least one U.S. address	ISI
Public address	Binary variable (1/0) denoting at least one public address	ISI
Private address	Binary variable (1/0) denoting at least one private address	ISI
Impact factor	Impact factor for journal in which paper is published	ISI/ Journal Citation Report
Patent Characteristics		
Patent in force	Binary variable (1/0) set to 1 if citation is received in years after patent grant	USPTO
Patent window	Binary variable (1/0) set to 1 if citation is received in year of patent grant	USPTO
Patent grant lag	Number of years between patent application and grant	USPTO
Patent scope	Number of national patent classes	USPTO
Number of claims	Number of claims in the patent	USPTO
Number of inventors	Number of inventors appearing on patent	USPTO
Number of assignees	Number of assignees appearing on patent	USPTO
Public assignee	Binary variable (1/0) denoting at least one public assignee	USPTO
All public assignee	Binary variable (1/0) denoting all public assignee	USPTO
Private assignee	Binary variable (1/0) denoting at least one private assignee	USPTO
All private assignee	Binary variable (1/0) denoting all private assignee	USPTO
U.S. assignee	Binary variable (1/0) denoting at least one U.S. based assignee	USPTO
Patent-Gene Characteristics		
Gene patents	Count of the number of patents for any given gene	USPTO/ Jensen & Murray 2005
Gene fragmentation (herfgene)	Herfindahl measure of concentration of ownership for a given gene using assignees on list for gene patents	USPTO/ Jensen & Murray 2005
OMIM gene	Binary variable (1/0) set to 1 if gene is listed in OMIM	OMIM
Cancer gene	Binary variable (1/0) set to 1 if gene is listed in Wellcome Cancer Gene Census	Wellcome Trust
Disease gene	Binary variable (1/0) set to 1 if gene is OMIM OR Cancer	OMIM/ Wellcome Trust

TABLE 2
Summary Statistics of Key Variables for Publications and Patents

Citation-Year Characteristics					
Key Variables	**n**	**Mean**	**s.d.**	**Min**	**Max**
Annual cite	12830	13.34	25.04	0	294
Total cite	12830	147.97	258.51	0	2369
Citation year	12830	2001	3.48	1988	2006
Paper age	12830	4.90	3.48	0	18
Paper Characteristics					
Paper year	1279	1997	2.78	1988	2005
Number of authors	1279	7.28	4.57	1	63
Number of addresses	1279	2.72	2.00	1	16
U.S. address	1279	0.80	0.40	0	1
Public address	1279	0.84	0.36	0	1
Private address	1279	0.35	0.48	0	1
Impact factor	1279	9.75	8.59	1	33.46
Patent Characteristics					
Patent in force	12830	0.58	0.49	0	1
Patent window	12830	0.10	0.30	0	1
Patent application year	1279	1997	2.57	1990	2003
Patent grant year	1279	2000	2.66	1993	2005
Patent grant lag	1279	3.30	1.51	0	11
Patent scope	1279	6.20	2.99	1	25
Number of claims	1279	15.61	14.20	1	137
Number of inventors	1279	2.56	1.40	1	14
Number of assignees	1279	1.13	0.42	1	4
Public assignee	1279	0.58	0.49	0	1
All public assignee	1279	0.54	0.50	0	1
Private assignee	1279	0.46	0.50	0	1
All private assignee	1279	0.42	0.49	0	1
U.S. assignee	1279	0.79	0.41	0	1
Patent-Gene Characteristics					
Gene patent 1	1279	0.335	0.47	0	1
Gene patent 2 to 4	1279	0.523	0.50	0	1
Gene patent 5 to 7	1279	0.099	0.30	0	1
Gene patent 8 to 10	1279	0.025	0.16	0	1
Gene patent 11 to 20	1279	0.017	0.13	0	1
Gene (max) fragmentation index	1279	0.67	0.30	0.107	1
OMIM gene	1279	0.25	0.43	0	1
Cancer gene	1279	0.076	0.26	0	1
Disease gene	1279	0.27	0.44	0	1

TABLE 3
Correlations Matrix

Independent Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
1 Patent in force	1.00																											
2 Patent window	-0.38	1.00																										
3 Paper year	-0.07	0.09	1.00																									
4 Number of authors	-0.01	0.02	0.17	1.00																								
5 Number of addresses	-0.03	0.01	0.17	0.49	1.00																							
6 U.S. address	0.05	0.00	0.04	-0.01	0.10	1.00																						
7 Public address	-0.02	-0.01	-0.09	-0.01	0.27	-0.02	1.00																					
8 Private address	0.03	0.01	0.15	0.22	-0.08	0.06	-0.57	1.00																				
9 Impact factor	-0.02	-0.01	-0.10	0.15	0.15	0.10	0.12	-0.05	1.00																			
10 Patent application year	-0.16	0.08	0.88	0.11	0.16	-0.07	-0.03	0.05	-0.05	1.00																		
11 Patent grant year	-0.30	0.07	0.73	0.11	0.15	-0.07	-0.02	0.04	-0.02	0.83	1.00																	
12 Patent grant lag	-0.23	-0.01	-0.21	0.01	-0.01	-0.01	0.03	-0.02	0.05	-0.23	0.34	1.00																
13 Patent scope	-0.01	0.00	-0.02	0.00	-0.01	-0.03	0.01	-0.01	0.01	-0.01	0.00	0.02	1.00															
14 Number of claims	-0.02	0.00	0.00	0.05	0.03	0.05	0.01	0.03	0.07	-0.01	0.04	0.08	0.17	1.00														
15 Number of inventors	-0.01	0.00	0.00	0.30	0.12	-0.04	0.03	0.04	0.13	0.01	0.02	0.02	0.04	0.01	1.00													
16 Number of assignees Public	-0.01	0.00	-0.01	0.15	0.14	-0.03	0.11	0.01	0.09	0.01	0.02	0.03	0.03	0.08	0.22	1.00												
17 assignees Private	0.01	-0.01	-0.12	-0.09	0.18	0.26	0.49	-0.58	0.13	-0.11	-0.09	0.02	-0.05	-0.01	0.02	0.09	1.00											
18 assignees	-0.01	0.01	0.12	0.15	-0.13	-0.26	-0.45	0.62	-0.10	0.10	0.09	-0.01	0.07	0.05	0.04	0.16	-0.91	1.00										
19 U.S. assignee	0.05	0.00	-0.01	-0.05	0.05	0.83	0.01	0.01	0.10	-0.13	-0.12	0.01	-0.04	0.07	-0.05	0.03	0.32	-0.30	1.00									
20 Gene patent 1	-0.02	0.01	0.11	-0.07	0.05	-0.01	0.11	-0.10	-0.03	0.15	0.10	-0.08	0.02	-0.08	0.02	0.04	0.05	-0.05	-0.04	1.00								
21 Gene patent 2 to 4	0.02	-0.01	-0.03	0.05	-0.05	0.02	-0.09	0.08	-0.01	-0.07	-0.05	0.03	-0.02	0.01	-0.01	0.00	-0.04	0.05	0.01	-0.73	1.00							
22 Gene patent 5 to 7	0.00	-0.01	-0.07	0.02	0.01	-0.01	0.02	-0.05	0.03	-0.06	-0.04	0.04	0.01	0.08	-0.04	-0.04	0.05	-0.06	0.03	-0.24	-0.36	1.00						
23 Gene patent 8 to 10	0.00	0.00	-0.03	0.03	0.01	-0.04	-0.06	0.05	0.07	-0.03	-0.01	0.03	-0.03	0.01	0.05	0.00	-0.06	0.07	0.00	-0.11	-0.17	-0.06	1.00					
24 Gene patent 11 to 20	0.00	0.00	-0.06	-0.03	-0.04	0.04	-0.03	0.07	-0.01	-0.06	-0.04	0.03	-0.01	0.02	0.01	-0.04	-0.06	0.05	0.03	-0.10	-0.15	-0.05	-0.02	1.00				
25 Fragmentation index	0.01	0.01	0.09	-0.11	-0.03	0.02	0.07	-0.11	-0.03	0.09	0.02	-0.12	-0.01	-0.08	-0.10	-0.26	0.04	-0.12	0.00	0.66	-0.36	-0.27	-0.18	-0.15	1.00			
26 OMIM gene	-0.03	0.00	-0.09	0.10	0.13	0.01	0.06	-0.08	0.08	-0.06	0.01	0.11	-0.06	0.05	0.04	0.04	0.03	-0.04	-0.02	-0.07	-0.01	0.07	0.04	0.05	-0.11	1.00		
27 Cancer gene	-0.01	-0.01	-0.06	0.07	0.04	0.02	0.06	-0.04	0.02	-0.06	-0.01	0.08	-0.03	0.10	-0.01	0.09	0.03	-0.02	0.02	-0.06	0.01	0.04	-0.01	0.06	-0.12	0.30	1.00	
28 Disease gene	-0.03	0.00	-0.09	0.10	0.13	0.01	0.07	-0.09	0.08	-0.06	0.01	0.11	-0.06	0.08	0.04	0.06	0.05	-0.04	-0.01	-0.07	0.00	0.06	0.04	0.06	-0.11	0.94	0.47	1.00

TABLE 4
List of Top Ten Most Frequently Appearing Journals

Journal Name	Frequency of Appearance	Percentage Gene Papers (%)	Journal Impact Factor (2005)
Journal of Biological Chemistry	202	15.78	5.85
The Proceedings of the National Academy of Sciences USA	137	10.70	10.23
Genomics	61	4.77	3.18
Science	47	3.67	30.93
Nature	46	3.59	29.27
Cell	34	2.66	29.43
Nature Genetics	34	2.66	25.80
Cancer Research	31	2.42	7.62
Biochemical & Biophysical Research Communications	28	2.19	3
Molecular and Cellular Biology	27	2.11	7.09

TABLE 5
Distribution of Genes by Number of Times Patented

Times Patented	Frequency: Number of Genes (Full Patented Population)	Frequency: Number of Genes (Paired Genes Sample)
1	2,844	1,535
2	905	611
3	350	266
4	139	100
5	55	46
6	38	34
7	17	16
8	13	12
9	9	5
10	4	4
11	1	1
12	1	1
13	2	2
14	2	2
20	2	2
Total	4,382	2,637

TABLE 6
Top 20 Most Patented Genes

Gene Name	Gene Function	Disease Linked (OMIM or Cancer)	Number of Patents/ Times Patented	Year of First Patent Application	Public, Private or Joint First Patent	% Patents with one or more private assignee
BMP7	Bone morphogenetic protein 7	N	20	1991	Private	100
CDKN2A	Cyclin-dependent kinase inhibitor 2A	Y	20	1994	Public	40
BRCA1	Breast cancer 1, early onset	Y	14	1995	Public	86
CA9	Carbonic anhydrase IX	N	14	1992	Private	7
NRG1	Neuregulin 1	Y	13	1993	Joint	85
SHH	Sonic hedgehog homolog	N	13	1994	Public	15
LEPR	Leptin receptor	Y	12	1994	Private	100
RGS16	Regulator of G-protein signalling 16	N	11	1994	Private	45
CD40LG	CD40 ligand	Y	10	1994	Private	100
IHH	Indian hedgehog homolog	Y	10	1994	Public	10
IL1RN	Interleukin 1 receptor antagonist	Y	10	1995	Private	100
VEGFC	Vascular endothelial growth factor C	N	10	1995	Private	60
CACNA2D1	Calcium channel, voltage-dependent, alpha 2/delta subunit 1	N	9	1991	Public	89
IKBKB	Inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase beta	N	9	1997	Private	89
LTF	Lactotransferrin	N	9	1993	Public	33
PTGER2	Prostaglandin E receptor 2, 53kDa	N	9	1994	Private	22
TERC	Telomerase RNA component	Y	9	1994	Private	67
PLA2G7	Phospholipase A2, group VII	Y	9	1995	Private	100
KLK4	Kallikrein 4	N	9	1996	Private	100
PIK3R5	Phosphoinositide-3-kinase, regulatory subunit 5, p101	Y	9	1997	Private	100

TABLE 7
Impact of Patent in Force[a,b]

	NBRM: $DV = FC$					
	Coefficients reported as incidence rate ratios, IRR					
	[7-1] Baseline	[7-2] Marginal Effects	[7-3] Full	[7-4] Patent Scope Interaction	[7-5] Patent Claims Interaction	[7-6] Full with Interaction
Independent Variables						
Patent window		0.94 (0.05)[0.06]	0.99 (0.02)[0.06]	0.99 (0.02)[0.02]	0.99 (0.02) [1.49]	0.99 (0.02) [1.48]
Patent in force		0.83*** (0.04)[0.16]	0.95** (0.02)[0.06]	0.95** (0.02)[0.06]	0.95** (0.02) [0.06]	0.95** (0.02) [0.06]
Patent in force x (Scope-mean scope)				0.99*** (0.00)[0.00]		0.99*** (0.00) [0.00]
Patent in force x (# of claims - mean # of claims)					1.00 (0.00) [0.00]	1.00 (0.00) [0.00]
Control Variables						
Number of authors	1.03*** (0.00)[0.01]	1.03*** (0.00)[0.01]				
Number of addresses	1.03*** (0.01)[0.01]	1.03*** (0.01)[0.01]				
U.S. address	0.93 (0.03)[0.05]	0.95 (0.04)[0.05]				
Public address	0.87*** (0.03)[0.05]	0.87*** (0.03)[0.05]				
Impact factor	1.09*** (0.00)[0.00]	1.09*** (0.00)[0.00]				
Paper fixed effects			χ^2 110000***	χ^2 110000***	χ^2 110000***	χ^2 110000***
Paper age fixed effects	χ^2 1242***	χ^2 1201***	χ^2 2868***	χ^2 2860***	χ^2 2853***	χ^2 2862***
Citation year fixed effects	χ^2 214***	χ^2 144***	χ^2 520***	χ^2 511***	χ^2 508***	χ^2 510***
Regression Statistics						
Log-likelihood	-41141	-41125	-32134	-32126	-32134	-32125
Wald chi-square (p)	0.00	0.00	0.00	0.00	0.00	0.00
Number of observations	12830	12830	12830	12830	12830	12830

[a] Robust standard errors in round parentheses. [b] Adjusted standard errors in square parentheses.[20] *p<0.05; **p<0.01; ***p<0.001

---

[20] In addition to the robust standard errors in Tables 7 and 8, we also report in square parentheses the adjusted standard errors using the aggregation techniques discussed in Bertrand, Duflo and Mullainathan (2004: 267) to correct for the long post-patent time on a large number of papers using differences-in-differences estimates.

TABLE 8
Patent Assignees, Thicket, Fragmentation and Useful Genes[a, b]

	NBRM: *DV = FC* Coefficients reported as incidence rate ratios, IRR							
	[8-1] Public Assignee vs. No Public Assignee	[8-2] Private Assignee vs. No Private Assignee	[8-3] All Public Assignee vs. All Private Assignee	[8-4] Patent Thicket	[8-5] Patent-Gene Fragment-ation Index	[8-6] OMIM Gene vs. No OMIM Gene	[8-7] Cancer Gene vs. No Cancer Gene	[8-8] Disease Gene vs. No Disease Gene
Independent Variables								
Patent window	0.99 (0.02)[0.06]	0.99 (0.02)[1.48]	1.00 (0.02)[0.69]	0.99 (0.02)[1.48]	0.99 (0.02)[1.47]	0.99 (0.02)[0.06]	0.99 (0.02)[1.47]	0.99 (0.02)[1.49]
Patent in force x public assignee	0.97 (0.02)[0.06]							
Patent in force x no public assignee	0.92*** (0.02)[0.06]							
Patent in force x private assignee		0.91*** (0.02)[0.06]						
Patent in force x no private assignee		0.98 (0.02)[0.06]						
Patent in force x all public assignee			1.00 (0.02)[0.06]					
Patent in force x all private assignee			0.94** (0.02)[0.06]					
Patent in force x gene patent 1				0.93** (0.02)[0.06]				
Patent in force x gene patent 2 to 4				0.97 (0.02)[0.06]				
Patent in force x gene patent 5 to 7				0.91** (0.03)[0.06]				
Patent in force x gene patent 8 to 10				0.97 (0.05)[0.06]				
Patent in force x gene patent 11 to 20				1.13 (0.10)[0.06]				
Patent in force					0.95* (0.02)[0.06]			
Patent in force x (mean frag_index - frag_index)					0.93* (0.03)[0.04]			
Patent in force x OMIM gene						0.92*** (0.02)[0.06]		
Patent in force x no OMIM gene						0.97 (0.02)[0.06]		
Patent in force x cancer gene							0.89** (0.03)[0.06]	
Patent in force x no cancer gene							0.96* (0.02)[0.06]	
Patent in force x disease gene								0.94** (0.02)[0.06]
Patent in force x no disease gene								0.96 (0.02)[0.06]
Control Variables								
Paper fixed effects	$\chi^2$110000***	$\chi^2$110000***	$\chi^2$110000***	$\chi^2$110000***	$\chi^2$110000***	$\chi^2$110000***	$\chi^2$110000***	$\chi^2$110000***
Paper age fixed effects	$\chi^2$2867***	$\chi^2$2858***	$\chi^2$2816***	$\chi^2$2847***	$\chi^2$2862***	$\chi^2$3030***	$\chi^2$2877***	$\chi^2$3028***
Citation year fixed effects	$\chi^2$519***	$\chi^2$508***	$\chi^2$524***	$\chi^2$516***	$\chi^2$517***	$\chi^2$685***	$\chi^2$521***	$\chi^2$681***
Regression Statistics								
Log-likelihood	-32131	-32128	-32132	-32128	-32132	-32132	-32131	-32133
Wald chi-square (p)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Number of observations	12830	12830	12830	12830	12830	12830	12830	12830

[a] Robust standard errors in round parentheses. [b] Adjusted standard errors in square parentheses. *p<0.05; **p<0.01; ***p<0.001
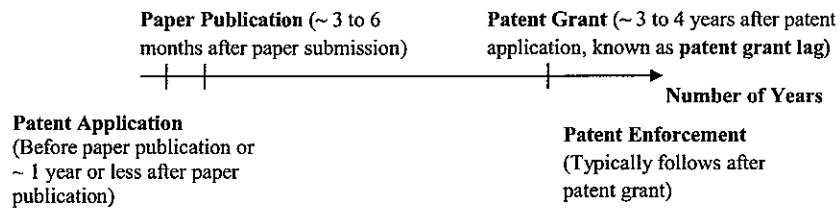
FIGURE 1
Timeline for a Typical Patent-Paper Pair

Paper Publication (~ 3 to 6 months after paper submission)

Patent Grant (~ 3 to 4 years after patent application, known as **patent grant lag**)

Number of Years

Patent Application
(Before paper publication or
~ 1 year or less after paper
publication)

Patent Enforcement
(Typically follows after
patent grant)

FIGURE 2
Distribution of Granted Gene Patents by Application Year
(full sample of 4273 gene patents)

Granted Gene Patents by Application Year