How Do Patents Affect Follow-On Innovation? Evidence from the Human Genome[†]

By Bhaven Sampat and Heidi L. Williams*

We investigate whether patents on human genes have affected follow-on scientific research and product development. Using administrative data on successful and unsuccessful patent applications submitted to the US Patent and Trademark Office, we link the exact gene sequences claimed in each application with data measuring follow-on scientific research and commercial investments. Using these data, we document novel evidence of selection into patenting: patented genes appear more valuable—prior to being patented—than non-patented genes. This evidence of selection motivates two quasi-experimental approaches, both of which suggest that on average gene patents have had no quantitatively important effect on follow-on innovation. (*JEL* I10, O31, O34)

Competitive markets may under-incentivize innovation due to the public good nature of new ideas. Intellectual property rights, such as patents, aim to address this under-investment problem by allowing inventors to capture a higher share of the social returns to their research investments. By awarding inventors a temporary right to exclude others from marketing their invention, patents aim to allow inventors

to earn quasi-rents (temporarily) as a way to recoup their research and development costs, thus providing dynamic incentives for investments in new technologies. Dating back at least to analyses such as Nordhaus (1969), optimal patent policy design has traditionally been framed as a trade-off between this benefit of providing incentives for the development of new technologies and the cost of deadweight loss from higher prices during the life of the patent.

Nordhaus-style models of optimal patent policy design have traditionally modeled innovations as isolated discoveries. However, in practice, many or most innovations are cumulative, in the sense that a given discovery may be an input into later follow-on discoveries. When innovation is cumulative, optimal patent policy design also depends on how patents on existing technologies affect follow-on innovation. One prominent example in recent policy debates is a patented human gene sequence. Sequenced genetic data are a research input into subsequent follow-on discoveries: by analyzing sequenced genetic data, scientists may discover links between genetic variations and diseases, and such knowledge can be applied to commercialize medical technologies such as pharmaceutical treatments and diagnostic tests. If patents on discoveries such as human genes affect follow-on innovation, that effect is a key additional input into optimal patent policy design. In this paper, we investigate whether patents on human genes have affected follow-on scientific research and product development. Our broad goal is to inform whether the Nordhaus-style trade-off between ex ante incentives and deadweight loss is sufficient for optimal patent policy design, or whether, at least in this context, the effects of patents on follow-on innovation need to be considered.

Investigating how patents on existing technologies affect follow-on innovation requires addressing two key challenges. First, in most markets it is difficult or impossible to take a given set of technologies which are claimed as intellectual property in patents and measure follow-on innovation on those technologies. Follow-on innovations which infringe the original patent legally require a license from the original innovator. But because licenses are very rarely publicly disclosed, constructing appropriate measures of follow-on innovation for any given patented invention is quite difficult. Second, a priori we expect a selection bias problem to arise if inventors are more likely to file for or to obtain patents on technologies that are inherently more valuable. This type of selection raises the concern that any measured differences in follow-on innovation across patented and non-patented technologies may reflect the selection of which technologies are patented, rather than a causal effect of patents on follow-on innovation.

The contribution of this paper is to construct new data and develop two new quasi-experimental approaches to address these challenges. To address the first challenge—measurement—we take advantage of the fact that US patent applications claiming human genes as intellectual property must disclose the exact DNA sequences claimed in the text of the patent. By applying bioinformatics methods (Jensen and Murray 2005), these DNA sequences can be annotated with gene identifiers, and these gene identifiers can in turn be linked to standard scientific and medical datasets providing measures of follow-on scientific research and product development related to each gene. Specifically, we measure the scientific publications related to each gene as an indicator of scientific research investments, and measure the use of genes in pharmaceutical clinical trials and diagnostic tests as

indicators of commercial research investments. Gene patents have been widely interpreted as sufficiently broad that these types of commercial follow-on inventions would require gene patent licenses. For example, then-United States Patent and Trademark Office (USPTO) biotechnology examination unit head John Doll noted in 1998: "…once a [gene] is patented, that patent extends to any use, even those that have not been disclosed in the patent" (Doll 1998, p. 690). This implies that our measures of follow-on innovation correspond closely to the way follow-on innovation has been defined in the theoretical literature on cumulative innovation, where licenses are required for follow-on innovations developed outside of the firm holding the patent.

Because we observe our measures of follow-on scientific research and product development for all genes, this DNA sequence-based linkage allows us to compare follow-on innovation across patented and non-patented genes. If patents were as good as randomly assigned across genes, this would be sufficient to estimate the causal effect of interest. However, if inventors are more likely to file for and obtain patents on technologies that are more valuable, then this type of simple comparison could instead reflect the selection of which genes are patented. By taking advantage of the fact that we observe our measures of follow-on innovation both before and after gene patents are granted, we are able to document novel evidence of selection into patenting: genes that will be patented in the future are the focus of more scientific research and more commercial investments prior to being patented, relative to genes that will not be patented. This evidence suggests that estimating a causal effect of patents on follow-on innovation requires constructing an empirical strategy to address this selection bias.

To address this second challenge (selection bias) we develop two new quasi-experimental methods for estimating how patents have affected follow-on innovation. First, we present a simple comparison of follow-on innovation across genes claimed in accepted versus rejected patent applications.[1] In our context, this method is reasonable if, conditional on being included in a patent application, whether a gene is granted a patent is as good as random. Again taking advantage of the fact that we observe our measures of follow-on innovation both before and after gene patents are granted, we document empirically that genes claimed in accepted and rejected patent applications are the focus of similar levels of scientific research and commercial investments prior to the applications being filed, providing evidence for the validity of this empirical approach. Second, we develop a novel instrumental variable for which patent applications are granted patents: the "leniency" of the assigned patent examiner. Patent examiners are charged with a uniform mandate: grant patents to patent-eligible, novel, non-obvious, and useful inventions. However, prior research has documented that in practice this mandate leaves patent examiners a fair amount of discretion (Cockburn, Kortum, and Stern 2003, Lemley and Sampat 2010, 2012). We leverage the interaction of this across-examiner heterogeneity with the quasi-random assignment of patent applications to examiners as a source of variation in which patent applications are granted patents. Past qualitative evidence, and

---

[1] Strictly speaking, patent applications are never formally rejected by the USPTO, only abandoned by applicants (see the discussion in Lemley and Sampat 2008). Online Appendix A provides both a qualitative discussion and some quantitative analyses of this issue. For brevity, we colloquially refer to such applications as "rejected" and discuss the relevance of this point for our measurement and empirical approaches as relevant throughout the text.

new empirical evidence we document, supports the assertion that the assignment of patent applications to examiners is plausibly random conditional on some covariates (such as application year and technology type), suggesting that the leniency of the patent examiner to which a patent application is assigned can provide a valid instrument for whether the patent application is granted a patent.

In contrast with what one would infer from a naïve comparison of follow-on innovation on patented and non-patented genes, both of our quasi-experimental approaches suggest that gene patents have not had quantitatively important effects on either follow-on scientific research or follow-on commercial investments. The estimates from our first quasi-experimental approach—comparing follow-on innovation across genes claimed in successful versus unsuccessful patent applications—document estimates which are economically small and meaningfully precise. While the estimates from our second quasi-experimental approach—using the leniency of the assigned patent examiner as an instrument for which patent applications are granted patents—are less precise, the fact that these two approaches generate similar conclusions provides additional confidence in our estimates.

When interpreting these empirical results through the lens of the well-developed theoretical literature on cumulative innovation (e.g., Green and Scotchmer 1995), several points are relevant to highlight. First, as best we can measure, very little of the follow-on innovation in this market is done by the gene patent holders, suggesting that follow-on innovation in this market will often require licensing agreements. Second, essentially all (96 percent) of the patent applications in our sample are assigned to for-profit firms, and in the vast majority of clinical trials (86 percent) both the patent assignee and the follow-on innovator are for-profit entities. Taken together, these facts suggest that cross-firm licensing contracts in this market seem to have operated at least somewhat efficiently. Consistent with this interpretation, there has been relatively little patent litigation observed in this market.

Interpreting these empirical results in the context of existing empirical evidence, perhaps most closely related to this paper is work by one of the authors (Williams 2013), who found that a non-patent form of database protection held by the private firm Celera on their version of the sequenced human genome was associated with large declines in follow-on scientific research and commercial product development, on the order of 30 percent. Why did Celera's intellectual property cause declines in follow-on innovation, while gene patents did not? Theoretical models tend to analyze stylized characterizations of intellectual property rights in which Celera's intellectual property could reasonably be seen as practically identical to patent protection. However, Celera's intellectual property differed from the gene patents we analyze in one key dimension. The sequenced genetic data in both the accepted and the rejected patent applications we analyze were disclosed in a way that enabled open access to the data for all prospective follow-on users. In contrast, Celera's data were disclosed in a much more restrictive way. We argue that this difference in disclosure may plausibly account for the different observed effects of Celera's intellectual property and gene patents on follow-on innovation outcomes. Importantly, this disclosure interpretation reinforces one important reason why our estimates may not generalize beyond the case of human gene patents: as we discuss in more detail in Section V, the USPTO's specific (and more stringent) requirements

for the disclosure of sequenced genetic data may have made the disclosure function particularly effective in our context.

From a policy perspective, one reason this disclosure feature of the patent system may be important is that it highlights a potential unintended consequence of a recent set of high-profile legal rulings on the case *Association for Molecular Pathology v. Myriad Genetics*.[2] The firm Myriad Genetics was granted patents on human genes correlated with risks of breast and ovarian cancer, and in June 2013 the US Supreme Court unanimously ruled to invalidate a subset of Myriad's gene patent claims, arguing that such patents "would 'tie up' … [genes] and … inhibit future innovation…" (p. 11). That is, the Court argued that gene patents had sufficiently strong negative effects on follow-on innovation that genes should be ineligible for patent protection. While, consistent with the Court's view, there has been widespread concern that patents on human genes may hinder follow-on innovation, as argued by a 2006 National Academies report and by Caulfield et al. (2013) there was no empirical evidence available to either support or refute that assertion prior to this paper. From the disclosure perspective discussed above, underlying this ruling seems to be an assumption that if genes are not patented, they would be placed in the public domain. But at least in the case of Celera, when Celera's patent applications were rejected the firm instead chose to rely on an alternative form of intellectual property which, taken at face value, resulted in both lower private returns and lower social returns to Celera's research investments. Stated differently, our analysis corresponds to a partial equilibrium analysis which holds disclosure constant, but in general equilibrium firms may choose to protect their inventions with, e.g., trade secrecy or other strategies rather than with patents if patent protection is not available. This type of behavioral response would be critical to assessing the potential welfare consequences of the US Supreme Court's ruling not just for human genes but also for other technologies which have recently been declared to be unpatentable.[3]

Methodologically, our two quasi-experimental approaches build on similar applications in labor economics and public finance. Our comparison of accepted and rejected patent applications builds on Bound (1989) and von Wachter, Song, and Manchester's (2011) investigations of the disincentive effects of disability insurance on labor supply, and Aizer et al.'s (2016) investigation of the effects of cash transfers on mortality. Likewise, our approach of using examiner "leniency" as a source of variation in which patent applications are granted patents builds on past work investigating the effects of incarceration length using variation across judges (Kling 2006), the effects of foster care using variation across foster care case workers (Doyle 2007, 2008), and the disincentive effects of disability insurance on labor supply using variation across disability insurance examiners (Maestas, Mullen, and Strand 2013). Especially given the relatively small number of patent law changes

---

[2] Online Appendix B provides some additional background on this case.

[3] More specifically, this same line of argument has shaped at least three other recent US Supreme Court rulings which have moved to restrict the set of discoveries eligible for patent protection in other industries (Kesselheim et al. 2013). First, in *Bilski v. Kappos* the Court invalidated patent claims on an investment strategy, announcing it supported a "high enough bar" on patenting abstract ideas that it would not "put a chill on creative endeavor and dynamic change." Second, in *Mayo v. Prometheus*, the Court invalidated patent claims on methods of using genetic variation to guide pharmaceutical dosing, expressing concern that "patent law not inhibit further discovery by improperly tying up the future of laws of nature." Finally, in *Alice Corp v. CLS Bank* the Court invalidated patent claims on software based on similar arguments.

in countries like the United States in recent years, these two new sources of quasi-experimental variation will likely provide valuable opportunities to investigate the effects of patents in a variety of other applications.[4]

Section I describes our data, and Section II documents some descriptive statistics on our data. Section III presents estimates from our first quasi-experimental approach, comparing follow-on innovation across genes claimed in successful and unsuccessful patent applications. Section IV presents estimates from our second quasi-experimental approach, using the leniency of the assigned patent examiner as an instrumental variable for whether the patent application was granted a patent. Section V provides some interpretations of our empirical results, and Section VI concludes.

I. Data

This section describes our data construction.[5] To fix ideas, we start by describing an example patent application (USPTO patent application 08/483,554) claiming intellectual property over the BRCA1 gene, and describe our measures of follow-on innovation in the context of that example (Section IA). We then describe in more detail how we construct our sample of USPTO patent applications claiming intellectual property over human genes (Section IB), and our gene-level measures of follow-on scientific research and product development (Section IC).

A. *Example: USPTO Patent Application 08/483,554*

On June 7, 1995, Myriad Genetics filed USPTO patent application number 08/483,554: *17Q-Linked Breast and Ovarian Cancer Susceptibility Gene*. This application was subsequently granted a patent on May 5, 1998 (US patent number 5,747,282), and would later become a focus of the US Supreme Court case *Association for Molecular Pathology v. Myriad Genetics*.

This patent claimed intellectual property over an isolated sequence of nucleotide bases (adenine, cytosine, guanine, and thymine), the sequence of which is listed explicitly in the patent (*SEQ ID NO:1: AGC TCG CTG…*). By comparing this sequence of nucleotide bases to the census of human gene sequences, we can uncover that this sequence corresponds to the BRCA1 gene.

The text of the Myriad patent describes how variation in the precise sequence of nucleotide bases in the BRCA1 gene can induce variation in an individual's risk of developing breast and ovarian cancers. For example, women with certain abnormal types of BRCA1 or BRCA2 genes have a 38 to 87 percent lifetime risk of developing breast cancer, relative to about 12 percent in the general population.[6]

Such links between genetic variation and diseases are referred to as genotype-phenotype links. In the case of the BRCA1 gene, scientific papers have investigated links between BRCA1 and breast cancer (181 publications), ovarian

---

[4]Indeed, a number of papers have already started to apply these methodologies in other contexts. See, for example, Gaule (2015); Feng and Jaravel (2018); Farre-Mensa, Hegde, and Ljungqvist (2017); and Kline et al. (2017).

[5]Online Appendix C describes our data construction in more detail.

[6]These statistics are drawn from GeneReviews, published by the US National Institutes of Health (NIH).

cancer (96 publications), and pancreatic cancer (3 publications).[7] Once scientific knowledge of a given genotype-phenotype link has been documented, this knowledge can be applied to develop medical technologies such as pharmaceutical treatments and gene-based diagnostic tests. In the case of the BRCA1 gene, 17 pharmaceutical clinical trials have been conducted that focus on mutations in the BRCA1 gene, and a BRCA1 genetic test is marketed by the firm Myriad Genetics.[8] The goal of our data construction is to trace these measures of follow-on scientific research and product development for each human gene, and to link the data to records of which human genes have been included in USPTO patent applications and granted patents.

B. *Constructing USPTO Patent Application Sample*

Our quasi-experimental approaches require constructing data on the census of published USPTO patent applications that claim intellectual property over human genes: both successful (granted patents) and unsuccessful (not granted patents). Traditionally, unsuccessful USPTO patent applications have not been made publicly available. However, as part of the American Inventors Protection Act of 1999, the vast majority of USPTO patent applications filed on or after November 29, 2000 are published in the public record—regardless of whether they are granted patents—at or before 18 months after the filing date.[9]

From the census of USPTO patent applications filed on or after November 29, 2000 and published on or before December 31, 2013, we identify the subset of applications claiming intellectual property over genes. To do this, we follow the methodology proposed by Jensen and Murray (2005), which can be briefly summarized as follows.[10] Since the early 1990s DNA sequences have been listed in USPTO patent applications in a standard format, labeled with the text "SEQ ID NO" (sequence identification number).[11] This standardized format allows for DNA sequences to be cleanly extracted from the full text of USPTO published patent applications. Once extracted, standard bioinformatics methods can be used to compare these sequences against the census of human gene DNA sequences in order to

---

[7] The data are drawn from the NIH Online Mendelian Inheritance in Man (OMIM) database, described below, and is accurate as of July 27, 2017.

[8] The data are drawn from the Citeline Pharmaprojects database (accurate as of July 9, 2012) and the NIH GeneTests.org database (accurate as of September 18, 2012), both described below.

[9] For more details, see http://www.uspto.gov/web/offices/pac/mpep/s1120.html and the discussion in Lemley and Sampat (2010). Most applications not published 18 months after filing are instead published 60 months after filing. Some US patent applications opt out of publication: Graham and Hegde (2013) document that around 8 percent of US applications opted for pre-grant secrecy of patent applications. For the NBER patent technology field "drugs and medical," which includes the most common patent class in our sample, the share was 3.5 percent.

[10] While we focus on the Jensen and Murray (2005) definition of gene patents, there are many different types of DNA-related patent claims: e.g., protein-encoding sequences, expressed sequence tags (ESTs), single-nucleotide polymorphisms (SNPs), sequence-based claims, and method claims that pertain to specific genes or sequences; see Scherer (2002) and Holman (2012) for more discussion. As one point of comparison, nearly all of the patent applications in our sample are included in the *DNA Patent Database* (http://dnapatents.georgetown.edu/SearchAlgorithm-Delphion-20030512.htm), which is constructed using a completely different methodology; we are grateful to Robert Cook-Deegan and Mark Hakkarinen for sharing the patent application numbers included in the *DNA Patent Database*, which enabled this comparison. Our empirical strategies could of course be applied to any empirically implementable definition of gene patent applications, but we are not aware of data on any other alternative definitions.

[11] See https://www.uspto.gov/web/offices/pac/mpep/s2422.html. For a legal perspective on disclosure of genetic innovations in patent applications more generally, see Berman and Schoenhard (2004).

annotate each sequence with standard gene identifiers that can in turn be linked to outside databases.[12]

We apply this Jensen and Murray (2005) methodology to construct two samples. First, we construct a "first-stage sample" that aims to include the census of published USPTO patent applications that claim any non-human (e.g., mouse) DNA sequences in their patent claims (Bacon et al. 2006). Second, we construct a "human gene sample" that includes only the subset of published USPTO patent applications that claim human genes (Lee et al. 2007). This human gene sample is the focus of our analysis, given the focus in this paper on investigating how patents on human genes have affected follow-on innovation. However, this sample includes far fewer patent applications than the first-stage sample: we have around 1,500 patent applications in the human gene sample, relative to around 14,000 patent applications in the first-stage sample. This matters for our second quasi-experimental design, for which it is (as we will discuss more below) econometrically useful to estimate variation in the examiner patent grant propensities in a separate sample of patent applications. Hence, we use the first-stage sample to estimate the examiner patent grant propensities, and apply those estimates to the examiners in our human gene sample.

It is worth noting that one limitation of our quasi-experimental approaches is that they can only be implemented over the time period when unsuccessful patent applications were published (that is, applications filed on or after November 29, 2000). Some criticisms of gene patents (such as Heller and Eisenberg 1998) focus on the subset of gene patents covering expressed sequence tags (ESTs), which Rai (2012) argues were less commonly granted in later years.[13]

C. *Measuring Follow-On Innovation*

We collect data on three measures of gene-level follow-on innovation: scientific publications as a measure of scientific research effort; and two measures of product commercialization: gene-related pharmaceutical research, and gene-based diagnostic tests.

We collect data on the scientific publications related to each gene from the Online Mendelian Inheritance in Man (OMIM) database, which catalogs scientific papers that have documented evidence for links between genetic variation and phenotypes. Using these data, we construct a count of the number of scientific papers published related to each gene—across all phenotypes—in each year.

Because of the long time lags between basic drug discovery and the marketing of new drugs, new approvals of drugs that take advantage of sequenced genetic data are just barely starting to enter the market.[14] Given these time lags, rather than using drug approvals as a measure of pharmaceutical research, we instead focus on an intermediate measure of drug discovery: namely, drug compounds under

---

[12] Following Jensen and Murray (2005), we focus attention on the subset of DNA sequences that are explicitly listed in the claims of the patent applications (excluding DNA sequences that are referenced in the text of the patent application—and hence listed in the SEQ ID NO format—but not explicitly claimed as intellectual property).

[13] For the purposes of our descriptive analyses we can (and do) apply the Jensen and Murray (2005) methodology to identify granted USPTO patents that claim human genes but were filed prior to November 29, 2000. The data are also drawn from Lee et al. (2007), and include USPTO patents granted through 2005.

[14] Nicholas Wade, "A Decade Later, Genetic Map Yields Few New Curves," *New York Times,* June 12, 2010.

development, as disclosed in clinical trials.[15] Specifically, we measure gene-related pharmaceutical clinical trials using the Citeline Pharmaprojects database, a privately-managed competitive intelligence database that tracks drug compounds in clinical trials and, critically for our project, assigns gene identifiers to compounds related to genetic variation on specific genes (see online Appendix C for details on how Citeline describes their construction of this measure). Using these data, we construct a count of the number of clinical trials related to each gene in each year.

Finally, we collect data on gene-based diagnostic tests from the GeneTests.org database, a genetic testing registry. Some gene-based tests provide individuals with information about disease risk, such as the BRCA tests related to risks of breast and ovarian cancers; other gene-based tests identify individuals as relatively more or less appropriate for a given medical treatment, such as a test which predicts heterogeneity in the side effects of the widely prescribed blood thinner warfarin. Using the GeneTests.org data, we construct an indicator for whether a gene is included in any gene-based diagnostic test as of 2012. Unfortunately, these data are only available in a cross section (not a panel).

A priori, the impact of patents on follow-on scientific research could differ from the impact of patents on product development. For example, many have argued that most patented inventions are made available to academic researchers on sufficiently favorable licensing terms that academics are able to continue their research (USPTO 2001). Hence, even if transaction costs hinder licensing agreements for commercial applications we could expect to see no impact of patents on measures of follow-on academic research such as scientific publications. Alternatively, patents may change researchers' incentives of whether to disclose the results of their research through academic publications (Moon 2011), in which case observed differences in scientific publications could be explained by differences in disclosure rather than by differences in the true amount of underlying scientific research. By contrast, we would not expect the product development outcomes we measure to be affected by such disclosure preferences given that the measures we observe are revealed in the natural course of firms commercializing and selling their technologies.

II. Descriptive Statistics

A. *Summary Statistics*

Table 1 presents patent application-level summary statistics on our first-stage sample and our human gene sample. Our first-stage sample includes 14,476 USPTO patent applications with DNA sequences listed in their claims, while our human gene sample includes 1,545 USPTO patent applications with human genes listed in their claims. As described in Section I, our sample includes US patent applications filed starting in 2000, given that the requirement for unsuccessful patent applications to be published came into force for patent applications filed on or after November 29, 2000. The first-stage sample ends with application year 2010. The human gene

---

[15] Data on clinical trial investments have also been used as a measure of research effort in prior work, starting with Acemoglu and Linn (2004) and Finkelstein (2004) and including more recently Galasso and Schankerman (2015).

TABLE 1—PATENT APPLICATION-LEVEL SUMMARY STATISTICS

	Mean	Standard deviation	Minimum	Maximum	Observations
Panel A. First-stage sample					
Application year	2005	2.715	2000	2010	14,476
0/1, patent granted as of 2010	0.3043	0.4601	0	1	14,476
Panel B. Human gene sample					
Application year	2002	0.692	2000	2005	1,545
0/1, patent granted as of 2010	0.2557	0.4364	0	1	1,545

Note: This table shows summary statistics for our patent application-level data in each of our two samples: panel A for the first-stage sample of patent applications, and panel B for the human gene sample of patent applications.

sample ends earlier, with application year 2005, because 2005 is the last year of data used as an input into the data construction done by Lee et al. (2007).[16]

In both samples, about 30 percent of the patent applications had been granted patents by 2010. This patent grant rate tabulated in Table 1 is a so-called simple grant rate that does not account for the fact that US patent applications can spawn closely related "new" applications, such as so-called continuations or divisionals. Carley, Hegde, and Marco (2015) use internal USPTO data to calculate simple and "family" (including patents granted to continuations or divisionals) grant rates in the universe of new utility patent applications filed at the USPTO from 1996–2005. In their sample, 55.8 percent of applications were granted patents directly, and including patents granted to children increases the allowance rate to 71.2 percent. For the NBER patent technology field "drugs and medical," which includes the most common patent class in our sample, the progenitor allowance rate is 42.8 percent and the family allowance rate is 60.7 percent. This measurement issue is relevant to our second empirical strategy—which analyzes application-level data—but not to our first empirical strategy, which analyzes gene-level data, and hence captures all patent applications related to each gene. In practice, this issue does not appear to be quantitatively important in our sample: the family grant rates in our first-stage and second-stage samples are 32.5 percent and 27.7 percent, which are relatively close to the analogous simple grant rates tabulated in Table 1 of 30.4 percent and 25.6 percent. As would be expected given these small differences, we show in online Appendix Table D.1 that if we implement our second empirical strategy using a family grant rate instead of a simple grant rate that our main conclusions are unchanged.

Online Appendix Table D.2 tabulates gene-level summary statistics on our follow-on innovation outcomes, for the sample of genes included in our human gene patent application sample ($N = 15,524$). The counts of scientific publications per gene in calendar year 2011 and clinical trials per gene in calendar year 2011 are both right-skewed: the median gene has a value of zero for each of these variables, but the maximum values are 22 and 230, respectively. Around 12 percent of genes are used in a diagnostic test as of calendar year 2012.

---

[16] Unfortunately, we do not know of a data source which applies the Lee et al. (2007) methodology to later years of data.

B. *Comparison of Patented and Non-Patented Human Genes*

Our interest in this paper is in comparing follow-on innovation across patented and non-patented genes. In this subsection, we investigate the selection process in order to analyze which genes are patented. On one hand, we may expect inventors to be more likely to both file for and be granted patents on technologies that are more valuable: inventors may be more willing to pay the time and monetary cost of filing patent applications for inventions that are more valuable, and patent applications claiming intellectual property over more valuable inventions may also be more likely to "clear the bar" and be granted patents. On the other hand, the USPTO grants patents based on criteria that may not closely correspond with measures of scientific and commercial value (Merges 1988). Hence, it is an empirical question whether inventions with higher levels of scientific and commercial value are more likely to be patented.

There is a remarkable absence of empirical evidence on this question, largely due to a measurement challenge: in most markets, it is very difficult to measure the census of inventions, and to link those inventions to patent records in order to identify which inventions are patented. The only other paper we are aware of which has undertaken such an exercise is Moser (2012), who constructs a dataset of innovations exhibited at world's fairs between 1851 and 1915 and documents that "high-quality" (award-winning) exhibits were more likely to be patented, relative to exhibits not receiving awards. Given this dearth of previous estimates, documenting evidence on the selection of inventions into patenting is itself of interest. In addition, if we do observe evidence of selection into patenting, that would imply that any measured differences in follow-on innovation across patented and non-patented genes may in part reflect the selection of which genes are included in patent applications and granted patents, as opposed to an effect of patents on follow-on innovation.

For this exercise, we start with the full sample of human genes ($N = 26{,}440$). As measured in our data, approximately 30 percent of human genes have sequences that were explicitly claimed in granted US patents ($N = 7{,}975$).[17] Figure 1 documents trends in follow-on innovation by year separately for genes that ever receive a patent (triangle-denoted solid blue series), and for genes that never receive a patent (circle-denoted dashed red series). In panel A of Figure 1, we plot the average log number of scientific publications by year in each year from 1970 to 2012.[18] In panel B, we plot the average log number of clinical trials by year in each year from 1995 to 2011.[19]

---

[17] As a point of comparison, Jensen and Murray (2005) document that as of 2005, approximately 20 percent of human genes had sequences that were explicitly claimed in granted patents. Because our sample includes patent applications that were granted patents after 2005, we would expect our estimate to be larger.

[18] We focus on the average log number of scientific publications by year because even within a calendar year, the number of publications per human gene is quite right-skewed. The pattern of selection that we document is unchanged if we instead plot the share of genes with at least one scientific publication by year (online Appendix Figure D.1, panel (a)), or the average number of scientific publications by year (online Appendix Figure D.1, panel (c)). Here and elsewhere, we add one to our outcome variables before logging them in order to include observations with no observed follow-on innovation.

[19] These years 1995 to 2011 are the only years for which the Pharmaprojects data are available. As with the scientific publications measure, we focus on the average log number of clinical trials by year because this variable is quite right-skewed. Again, the pattern of selection that we document is unchanged if we instead plot the share of genes with at least one clinical trial by year (online Appendix Figure D.1, panel (b)), or the average number of clinical trials by year (online Appendix Figure D.1, panel (d)).
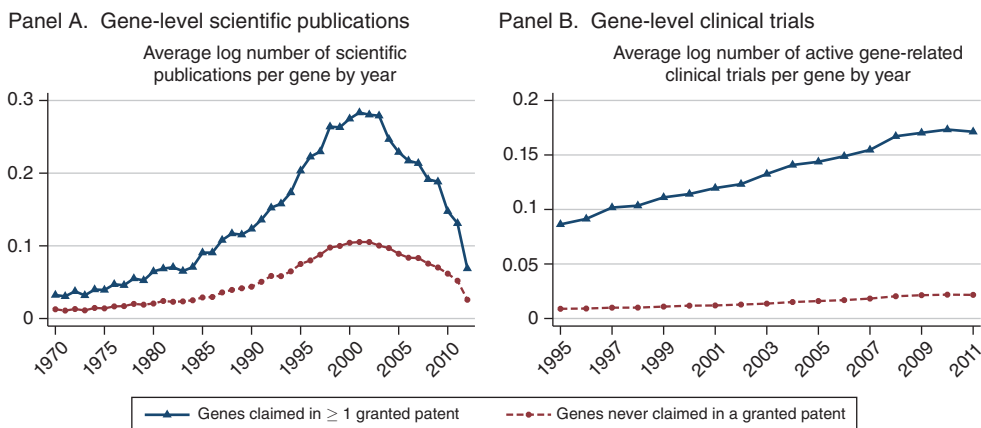
Panel A. Gene-level scientific publications   Panel B. Gene-level clinical trials

FIGURE 1. FOLLOW-ON INNOVATION ON PATENTED AND NON-PATENTED HUMAN GENES

Notes: This figure plots trends in follow-on innovation by year separately for genes that ever receive a patent and for genes that never receive a patent. The figure is constructed from gene-level data. Panel A uses gene-level scientific publications as a measure of follow-on innovation, and plots the average log number of scientific publications by year in each year from 1970 to 2012. Panel B uses gene-level clinical trials as a measure of follow-on innovation, and plots the average log number of clinical trials by year in each year from 1995 to 2011.

Because by construction our human gene patents are measured starting in the mid-1990s (when the SEQ ID NO notation was introduced), the cleanest test of selection into patenting is a comparison of pre-1990 follow-on innovation across (subsequently) patented and non-patented genes. While the clinical trials data only start later (in 1995), the scientific publications data are available prior to 1990 and therefore can be used for this comparison. Looking at the data series in Figure 1, panel A from 1970 to 1990 provides clear evidence of positive selection: genes that will later receive patents were more scientifically valuable—based on this publications measure—prior to being patented. Moreover, even within the 1970 to 1990 period this positive selection appears not just in the levels of follow-on innovation, but also in the trends: genes that will later receive patents appear to have divergent trends in scientific publications relative to genes that will never receive patents, even before any patents are granted.

These patterns—a level difference, and a divergence in trends—also appear in the clinical trials data presented in panel B of Figure 1, although it is not possible to cleanly separate selection and "treatment" (that is, any causal effect of patents on subsequent research effort) because that data series starts in 1995. Likewise, if we tabulate the probability that genes are used in diagnostic tests for patented and non-patented genes, 13.6 percent of patented genes are used in diagnostic tests, compared to 6.2 percent of non-patented genes.

Taken together, these patterns suggest strong evidence of positive selection: genes that will later receive patents appear more scientifically and commercially valuable prior to being granted patents, relative to genes that will never receive patents. This evidence is important for three reasons. First, this evidence suggests that our measures of value (pre-patent filing scientific publications and clinical

trials) are correlated with patenting activity, which supports the idea that these measures can provide a meaningful basis for assessing selection in our two quasi-experimental approaches. Second, this analysis provides novel evidence on the selection of technologies into patenting, in the spirit of Moser (2012). Third, this evidence implies that a simple comparison of follow-on innovation across patented and non-patented genes is unlikely to isolate an unbiased estimate of how gene patents affect follow-on innovation. While a naïve comparison that did not account for selection would conclude based on Figure 1 that patents encourage follow-on innovation, our two quasi-experimental approaches will suggest a different conclusion.

III. Comparison of Accepted and Rejected Patent Applications

Our first quasi-experimental source of variation investigates a simple idea, which is whether genes that were included in unsuccessful patent applications can serve as a valid comparison group for genes that were granted patents.

This approach has previously been applied to other research questions in labor economics and public finance. For a wide range of public programs, a key challenge in identifying the effect of the program is identifying a plausible counterfactual: what would recipients' lives have been like in the absence of the program? Dating back at least to the work of Bound (1989), researchers have proposed constructing this counterfactual by comparing accepted and rejected applicants. The validity of this approach of course depends on the extent to which accepted and rejected applicants differ on observable and unobservable characteristics, and researchers have hence explored the validity of this approach by comparing accepted and rejected applicants based on observable characteristics fixed at the time of application. For example, Aizer et al.'s (2016) analysis of accepted and rejected applicants for the Mothers' Pension program documented balance tests comparing accepted and rejected applicants on various covariates fixed at the time of application, such as mother's age and family size. If accepted and rejected applicants are similar on observables, then under the assumption that they are also similar on unobservables rejected applicants can provide a plausible counterfactual for the outcomes of accepted applicants in the absence of the program.[20]

We apply an analogous framework to our dataset of gene patent applications, in which some applications are accepted and some are rejected. One nuance is that our outcomes of interest are at the gene level, not at the patent application level. If every patent application included only one gene, and each gene was never included in more than one patent application, this distinction would be irrelevant. But in practice, patent applications can include more than one gene, and a given gene can be included in more than one patent application. To address this divergence from the simplest thought experiment, we divide the sample of genes that are ever included

---

[20] Note that Bound (1989) and others (including Aizer et al. 2016) argue that in their context they can make a slightly different assumption, which is that rejected applicants are better in terms of counterfactual outcomes than are accepted applicants. For example, in his context of disability insurance, Bound (1989) argues that rejected applicants are healthier and more capable of work, and thus that their labor force participation should provide an upper bound for what work could be expected of accepted beneficiaries. In our view, this type of "signing the bias" argument does not have a natural analog in our empirical context.

in any patent application into two subsamples: a subsample of genes that is included in at least one accepted patent application, and a subsample of genes included in at least one patent application but never in an accepted patent application.

Defined in this way, a comparison of follow-on innovation on genes included in accepted and rejected patent applications will be valid if—conditional on being included in a patent application—whether a gene is granted a patent is as good as random. A priori, it is not clear that this offers a valid comparison. The USPTO is responsible for assessing whether patent applications should be granted patents based on five criteria: patent-eligibility (35 USC. §101), novelty (35 USC. §102), non-obviousness (35 USC. §103), usefulness (35 USC. §101), and the text of the application satisfying the disclosure requirement (35 USC. §112). These criteria suggest that patent grants are not quasi-randomly assigned across patent applications.[21] However, because any given patent application in our sample may claim intellectual property rights over more than one gene, and given that our interest is in how patent grants affect gene-level outcomes, what is required for this first quasi-experimental approach to be valid is that—conditional on being included in a patent application—whether a gene is granted a patent is as good as randomly assigned. That is, the relevant assumption is that genes that are included in patent applications that are granted patents are comparable to genes that are included in patent applications that are not granted patents. Analogous to the balance tests used in the previous literature to assess the validity of accepted/rejected designs, we will document a comparison of "accepted" and "rejected" genes based on covariates fixed at the time of application (in our case, gene-level scientific publications and gene-level clinical trial activity at the time of application). As we will see, "accepted" and "rejected" genes look quite similar based on these observables at the time the patent application is filed. Hence, we will conclude that "rejected" genes can provide a plausible counterfactual for what follow-on innovation on the "accepted" genes would have looked like in the absence of being granted patents. While this comparison is quite simple, we will see that the resulting estimates are similar to the estimates from our second quasi-experimental source of variation (the examiner leniency variation, presented in Section IV).

A. *Graphical Analysis*

For this exercise, we start with the sample of human genes included in at least one patent application in our USPTO patent applications sample ($N = 15,524$; 59 percent of the full sample of 26,440 human genes). Of this sample, 4,858 genes are claimed in a patent application that is subsequently granted a patent (31 percent), relative to 10,666 genes that are never observed to be subsequently granted a patent (69 percent).[22] Panel A of Figure 2 illustrates the time pattern of when the patented

---

[21] Although, to foreshadow our second empirical approach, note that if patent applications were randomly assigned to patent examiners and patent examiners—who varied in their leniency—completely determined grant outcomes, that would be one substantive example of why patent grants could be as good as randomly assigned across accepted and rejected applications.

[22] Note that this 4,858 figure is lower than the 7,975 figure in Section II, because we here focus only on patents granted on patent applications filed after November 29, 2000 (the date when unsuccessful applications began to be published).
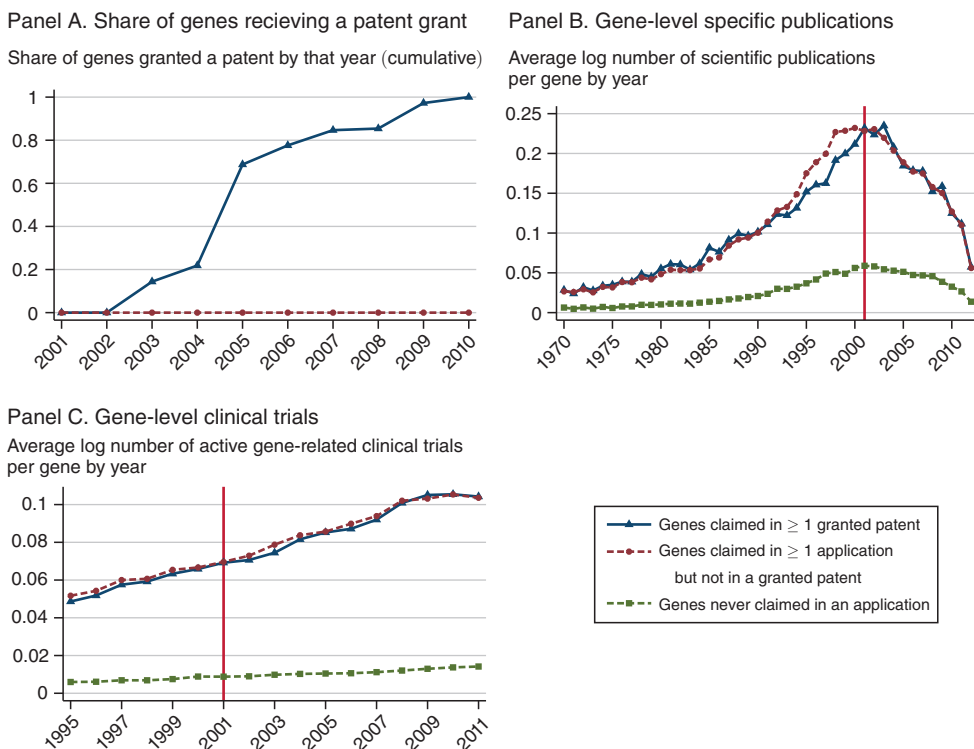
Panel A. Share of genes recieving a patent grant

Share of genes granted a patent by that year (cumulative)

Panel B. Gene-level specific publications

Average log number of scientific publications per gene by year

Panel C. Gene-level clinical trials

Average log number of active gene-related clinical trials per gene by year

Legend:
- Genes claimed in ≥ 1 granted patent
- Genes claimed in ≥ 1 application but not in a granted patent
- Genes never claimed in an application

FIGURE 2. PATENTS AND FOLLOW-ON INNOVATION ON HUMAN GENES CLAIMED IN ACCEPTED/REJECTED PATENT APPLICATIONS

Notes: This figure plots trends in patenting and follow-on innovation by year separately for three groups of genes: genes claimed in at least one granted patent; genes claimed in at least one patent application but never in a granted patent; and (in panels B and C) genes never claimed in a patent application. The figure is constructed from gene-level data. Panel A documents the share of genes receiving a patent grant by year; by construction, this is zero for the circle-denoted red dashed line in all years and reaches one for the triangle-denoted blue line in 2010; the intermediate years simply illustrate the time path of patent grants between 2001 and 2010 for the triangle-denoted blue line. Panel B uses gene-level scientific publications as a measure of follow-on innovation and plots the average log number of scientific publications by year in each year from 1970 to 2012. Panel C uses gene-level clinical trials as a measure of follow-on innovation and plots the average log number of clinical trials by year in each year from 1995 to 2011. The vertical line in the calendar year 2001 in panels B and C denotes that, because this figure focuses on patents that were filed in or after November 2000, all years prior to 2001 can be considered a pre-period and used to estimate the selection of genes into patenting based on pre-patent filing measures of scientific research (publications) and commercialization (clinical trials).

group receives its (first) patent grant over time. Over half of these genes have received a patent by 2005, and (by construction) all have received a patent by 2010.

Panels B and C of Figure 2 document trends in follow-on innovation by year. As in Figure 1, we plot the average log number of scientific publications by year in each year from 1970 to 2012 (Figure 2, panel B), and the average log number of clinical trials by year in each year from 1995 to 2011 (Figure 2, panel C).[23] The solid blue

---

[23] Online Appendix Figure D.2 documents analogous figures if we instead plot the share of genes with at least one scientific publication or clinical trial by year (panels (a) and (b)) or the average number of scientific publications or clinical trials by year (panels (c) and (d)).

triangle-denoted line represents genes claimed in at least one granted patent, and the dashed red circle-denoted line represents genes claimed in at least one patent application but never in a granted patent.

As a point of comparison, the dashed green square-denoted line represents genes never claimed in a patent application filed after November 29, 2000 ($N = 10,916$; 41 percent of the full sample of 26,440 human genes). Comparing this group of genes to the other two groups of genes, we see clear evidence of selection into patent filing: genes included in successful and unsuccessful patent applications are much more valuable both scientifically (publications) and commercially (clinical trials) prior to the patent application filing compared to genes that are never claimed in a patent application. In terms of interpreting the results in Section IIIB, the data suggest that the major source of selection into which genes are patented is selection in which genes are included in patent applications (as opposed to which genes are granted patents, conditional on being included in patent applications).

The key comparison of interest in this section is across the first two data series: genes included in successful patent applications, and genes included in unsuccessful patent applications. The vertical line in calendar year 2001 denotes that because this figure focuses on patent applications filed in or after November 2000, all years prior to 2001 can be considered a "pre-period" and used to estimate the selection of genes into patenting based on pre-patent filing measures of follow-on innovation. Strikingly, we see little evidence of selection in pre-2001 levels or trends of our two follow-on innovation measures once we limit the sample to genes included in patent applications. For scientific publications (panel B of Figure 2), the two groups follow each other quite closely from 1970 to 1990, and diverge slightly in trends from 1990 to 2000—with genes that will subsequently be included in unsuccessful patent applications having slightly more scientific publications.[24] For clinical trials (panel C of Figure 2), the two groups follow each other quite closely in both levels and trends over all available pre-2001 years of data. Taken at face value, the similarity of these two groups in pre-2001 outcomes provides evidence for the validity of this empirical approach. A priori, one might have expected genes that were more scientifically or commercially valuable to have been more likely to receive patents. However, conditional on being included in a patent application, this appears not to be the case.

Looking at the post-2001 time period, we see that although these two groups of genes diverge (by construction) in whether they are claimed in granted patents (panel A of Figure 2), we do not see any evidence of a divergence in follow-on innovation outcomes. That is, these figures suggest that gene patents have not had a quantitatively important effect on either follow-on scientific research or on follow-on commercialization.

B. Regression Analysis

We quantify the magnitudes of these patterns in a regression framework in Table 2. Because our scientific publication and clinical trial outcomes are quite skewed, a proportional model or binary outcome (measuring "any follow-on innovation") is

---

[24] Note that this slight divergence is not apparent in the robustness check documented in online Appendix Figure D.2, panel (c), where we plot the average number of scientific publications by year.

TABLE 2—PATENTS AND FOLLOW-ON INNOVATION ON HUMAN GENES CLAIMED IN
ACCEPTED/REJECTED PATENT APPLICATIONS: REGRESSION ESTIMATES

	log of follow-on innovation in 2011–2012	Any follow-on innovation in 2011–2012
	(1)	(2)
Panel A. Scientific publications		
Patent granted	0.0019	−0.0014
	(0.0060)	(0.0054)
Mean of dependent variable	0.1104	0.1094
Observations	15,524	15,524
Panel B. Clinical trials		
Patent granted	0.0006	−0.0015
	(0.0080)	(0.0043)
Mean of dependent variable	0.1038	0.0659
Observations	15,524	15,524
Panel C. Diagnostic test		
Patent granted	–	−0.0092
	–	(0.0056)
Mean of dependent variable	–	0.1199
Observations	–	15,524

Notes: This table estimates differences in follow-on innovation on genes claimed in at least one granted patent relative to genes claimed in at least one patent application but never in a granted patent. The sample for these regressions is constructed from gene-level data, and includes genes claimed in at least one patent application in our USPTO human gene patent application sample ($N = 15{,}524$). Each coefficient is from a separate regression. Estimates are from ordinary-least-squares models. Heteroskedasticity robust standard errors.

more econometrically appropriate than modeling the outcome in levels. We focus on the log of follow-on innovation and (separately) an indicator for any follow-on innovation.

Given the absence of strong visual evidence for a difference in follow-on innovation across patented and non-patented genes, our focus here is on what magnitudes of effects can be ruled out by our confidence intervals. Across these specifications, our 95 percent confidence intervals tend to reject declines or increases in follow-on innovation on the order of more than 5–15 percent. For brevity, we focus on interpreting the log coefficients. For our measures of follow-on scientific research (publications; panel A of Table 2) and commercialization (clinical trials; panel B of Table 2), the 95 percent confidence intervals can reject declines or increases of more than 2 percent. For our measure of diagnostic test availability (only measured as a binary indicator; panel C of Table 2), we estimate that genes receiving patents had a 0.9 percentage point decrease in the likelihood of being included in a diagnostic test as of 2012 relative to genes included in patent applications but not granted patents (statistically significant at the 10 percent level). Our 95 percent confidence interval can reject declines of greater than 2 percentage points and reject increases of more than 0.2 percentage points. Relative to a mean of 12 percent, this confidence interval suggests that we can reject declines in this outcome of greater than 17 percent.[25]

_____

[25] As a point of comparison, only 3 percent of genes never included in a patent application are included in a diagnostic test as of 2012.

IV. Analyzing Examiner-Level Variation in Patent Grant Propensities

Our second source of quasi-experimental variation constructs an instrumental variables strategy for predicting which patent applications are granted patents. Our key idea is to build on previous research which has established that although patent examiners are charged with a uniform mandate, in practice examiners have a fair amount of discretion, and this discretion appears to translate into substantial variation in the decisions different examiners make on otherwise similar patent applications (Cockburn, Kortum, and Stern 2003; Lichtman 2004; Lemley and Sampat 2010, 2012).[26] In the spirit of prior analyses such as Kling (2006), we leverage these patterns in order to use variation in the "leniency" of different patent examiners as a predictor of which patent applications are granted patents.

The exclusion restriction for this instrumental variables approach requires assuming that the examiner only affects follow-on innovation through the likelihood that a gene is patented. As we describe below, the institutional context suggests that the assignment of patent applications to USPTO patent examiners should be effectively random conditional on some covariates (such as application year and technology type). While the exclusion restriction is inherently untestable, we will document empirically that—consistent with our qualitative description of the institutional context—genes assigned to "lenient" and "strict" examiners look similar on observable characteristics fixed at the time of patent application.[27]

To motivate our empirical specification, Section IVA provides some qualitative background on the key institutional features underlying our empirical strategy.[28]

A. *Assignment of Patent Applications to Patent Examiners*

A central USPTO office assigns application numbers to incoming patent applications, as well as patent class and subclass codes detailing the type of technology embodied in the application.[29] These class and subclass numbers determine which of approximately 300 so-called Art Units—specialized groups of examiners—will review the application.[30] Within an Art Unit, a supervisory patent examiner assigns the application to a patent examiner for review. While the patent application process

---

[26] One of the individuals interviewed by Cockburn, Kortum, and Stern (2003, p. 21) described this variation informally by saying "there may be as many patent offices as there are patent examiners." Similarly, a trade publication written by a former USPTO patent examiner and current patent agent (Wolinsky 2002, p. 1) described this variation by saying "The successful prosecution of a patent application at the USPTO requires not only a novel invention and adequate prosecution skills, but a bit of luck… If you knew the allowance rate of your examiner, you could probably estimate the odds of getting your patent application allowed."

[27] While conditional random assignment of applications to examiners assuages many potential concerns about this exclusion restriction, some additional issues remain. In particular, while we focus on variation in patent grant propensity, examiner heterogeneity may also manifest itself in other ways, such as the breadth of patent grants (in terms of the number or strength of allowed claims) and time lags in grant decisions (Cockburn, Kortum, and Stern 2003).

[28] The discussion in Section IVA draws heavily on Cockburn, Kortum, and Stern (2003); US Government Accountability Office (2005); Lemley and Sampat (2012); and Frakes and Wasserman (2017). See online Appendix A for more detail on the USPTO patent examination process.

[29] There are currently over 450 patent classes, and more than 150,000 subclasses; see http://www.uspto.gov/patents/resources/classification/overview.pdf.

[30] See http://www.uspto.gov/patents/resources/classification/art/index.jsp. For the current version of the class/subclass-to-Art Unit concordance, see http://www.uspto.gov/patents/resources/classification/caau.pdf. The main Art Units in our sample are from the 1600 group (Biotechnology and Organic Chemistry).

up to this point is quite structured, from this point forward substantial discretion is left in the hands of individual examiners, who are responsible for determining which (if any) claims in the application are patentable.

Because no "standard" method for the within-Art Unit assignment of applications to examiners is uniformly applied in all Art Units, Lemley and Sampat (2012) conducted written interviews with roughly two dozen current and former USPTO examiners to inquire about the assignment process. While the results of these interviews suggested that there is not a single "standard" assignment procedure that is uniformly applied in all Art Units, these interviews revealed no evidence of deliberate selection or assignment of applications to examiners on the basis of characteristics of applications other than those observed in standard USPTO datasets (on which we can condition). For example, in some Art Units supervisors reported assigning applications to examiners based on the last digit of the application number; because application numbers are assigned sequentially in the central USPTO office, this assignment system, while not purposefully random, would be functionally equivalent to random assignment for the purposes of this study. In other Art Units, supervisors reported placing the applications on master dockets based on patent classes and subclasses, with examiners specializing in those classes (or subclasses) being automatically assigned the oldest application from the relevant pool when requesting a new application. Our key conclusion from this institutional context is that effective conditional random assignment of applications to examiners—within Art Unit and application year—is plausible. Consistent with this assumption, we will document that patent applications assigned to "lenient" and "strict"'examiners look similar based on observable characteristics fixed at the time of patent application.

From a practical perspective, it is worth noting that informational barriers limit the extent to which we would expect patent applications to be systematically sorted across examiners in a way that would be problematic for our empirical specifications. Because of the limited attention given to patents prior to their assignment to a specific examiner, and the judgment required to determine the characteristics of a given invention, it seems plausible that informational barriers would impose real constraints on sorting (this argument has been made in more detail by Merges 1999).[31] In particular, the "patentability" of applications is difficult to assess ex ante, and there is no evidence that supervisory patent examiners attempt to do so before assigning applications to particular examiners.

B. *Examiner Leniency Variation*

To the best of our knowledge, Kling (2006) was the first paper to leverage this type of variation: in his case, using the random assignment of court cases to judges as an instrument for incarceration length. He adopted the jackknife instrumental variables (JIVE1) approach proposed by Angrist, Imbens, and Krueger (1999),

---

[31] Merges (1999) also argues that although sorting of patent applications to examiners may be efficient, an additional barrier to such sorting is the strong "all patents are created equal" tradition at the USPTO, which cuts strongly against any mechanism for separating and sorting patents.

which predicts the judge effect for each case based on data for all other cases.[32] However, JIVE estimators have been criticized for having undesirable finite sample properties (see, e.g., Davidson and MacKinnon 2006). In our case, a natural alternative is to adopt a two-sample two-stage least squares (TS2SLS) variant of Angrist and Krueger's (1992) two-sample instrumental variable estimator. Specifically, for each patent examiner in our sample, we observe the decisions that examiner makes on non-human gene patent applications, and we can use that separate sample to estimate variation in leniency across examiners.[33] Motivated by the institutional context, we condition on Art Unit-by-application year fixed effects, so that we capture an examiner's patent grant propensity relative to other examiners reviewing applications in that Art Unit in that application year.[34] We use the two-sample two-stage least squares standard error correction provided by Inoue and Solon (2010), and cluster at the patent application level by extending Pacini and Windmeijer's (2016) generalization to the heteroskedastic case.[35]

C. *First-Stage Estimates*

Figure 3 provides a visual representation of our first stage. In our first-stage sample, we calculate the mean grant rate for each examiner, residualized by Art Unit-by-application year fixed effects, and relate this measure of examiner "leniency" to patent grant outcomes.[36] Visually, there is a strong relationship.

To quantify this relationship, we estimate the following equation for a patent application i examined by patent examiner j filed in year t assigned to Art Unit a in our first-stage sample:

$$(1) \qquad \mathbf{1}\left(\text{patent grant}\right)_{ijta} = \alpha + \beta \cdot \mathbf{Z}_{ijta} + \Sigma_{ta}\mathbf{1}\left(\text{art unit}\right)_{ta} + \varepsilon_{ijta},$$

where the outcome variable $\mathbf{1}\left(\text{patent grant}\right)_{ijta}$ is an indicator variable equal to 1 if patent application i was granted a patent, $\mathbf{Z}_{ijta}$ is the examiner's mean non-human gene patent grant rate, and $\Sigma_{ta}\mathbf{1}\left(\text{art unit}\right)_{ta}$ are a set of Art Unit-by-application year fixed effects.

In our first-stage sample, we estimate a β coefficient in this specification of 0.876, with a standard error of 0.037. This point estimate implies that a 10 percentage point increase in an examiner's average patent grant rate is associated with a 8.8 percentage point increase in the likelihood that a patent application is granted a

---

[32] Much of the subsequent literature (e.g., Doyle 2007) has approximated JIVE through an informal leave out mean approach.

[33] In practice, our results are similar if we use a leave-one-out-mean approach that computes examiner leniency on the same sample as our outcomes are measured; see online Appendix Tables D.3 and D.4.

[34] A separate concern is that if we only observed a small number of applications per examiner, our estimate of the variation in leniency across examiners would be overstated. To address this concern, we limit the first-stage sample to examiners that saw at least ten applications (Heckman 1981, Greene 2001). To match our conceptual thought experiment, we also limit the sample to Art Unit-years with at least two examiners.

[35] Conceptually, we would prefer to cluster our standard errors at the gene level. However, only a very small share of the patent applications in our first-stage sample can be matched to gene identifiers, making that adjustment infeasible. As a second best, we instead cluster at the patent application level in our two-sample two-stage least squares specifications, and cluster at the gene level in robustness checks where that is feasible, such as in the leave-one-out-mean approach in online Appendix Tables D.3 and D.4.

[36] We will describe the overlaid plot of lighter yellow triangles in Section IVD.
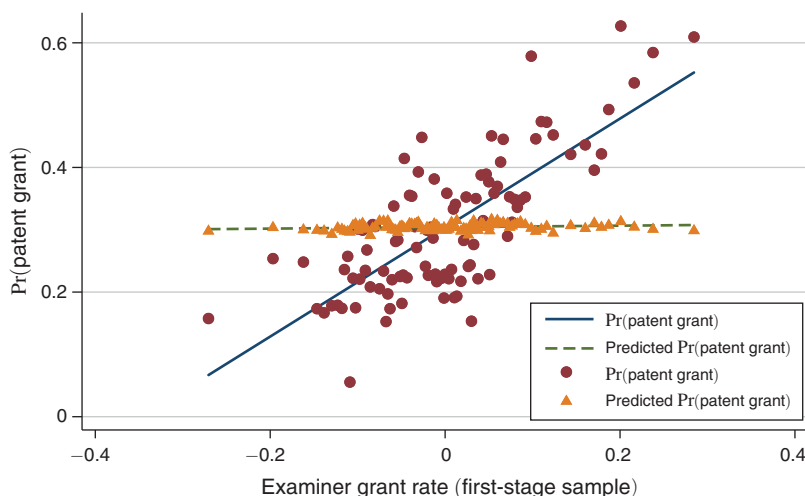
FIGURE 3. PROBABILITY OF PATENT GRANT BY EXAMINER LENIENCY

Notes: The figure relates our examiner leniency measure, residualized by Art Unit-by-application year fixed effects, to two variables: (1) the patent grant rate and (2) the predicted patent grant rate, where we predict patent grant as a function of our two measures of patent value fixed at the time of application (patent family size and claims count). All measures are constructed in our first-stage sample ($N = 14{,}476$).

patent.[37] The *F*-statistic on the examiner's grant rate is on the order of 500, well above the rule of thumb for weak instruments (Stock, Wright, and Yogo 2002).

As a robustness check, in online Appendix Table D.5 we replace Art Unit-by-year fixed effects in this specification with Art Unit-by-year-by-class-by-subclass fixed effects, on the subsample for which these finer fixed effects can be estimated. The point estimates from this more stringent specification are very similar to and not statistically distinguishable from the baseline point estimate described above, suggesting that at least in our context, variation in the measured leniency of different examiners is unlikely to be generated by the systematic sorting of patent applications by classes or subclasses.[38]

---

[37] Our patent grant outcome is measured as of 2010 and is censored for patent applications that are still in the process of being examined, but this censoring should be less of a concern for earlier cohorts of gene patent applications. The point estimates on a subsample of early cohorts of applications are very similar to our baseline point estimates (results not shown), suggesting that censoring appears to not substantively affect the magnitude of the estimated first-stage coefficient. Note that this is likely because the Art Unit-by-application year fixed effects largely account for differences in the probability of patent grant that are mechanically related to time since application. Given this similarity, we retain all cohorts of gene patent applications to retain a larger sample size.

[38] Righi and Simcoe (2017) provide a set of formal tests for random assignment of patent applications to examiners, and show that random assignment is rejected in the full sample of patent applications. Specifically, they document that in some art units examiners appear to specialize in reviewing particular types of technologies. This fact is consistent with the qualitative literature discussed above (such as Lemley and Sampat 2012), which suggested that in some art units examiners are responsible for reviewing particular classes/subclasses and technology types. In general, this type of technology specialization based on observables could raise the concern that there is also technology specialization based on *unobservables* that are correlated with both patentability and our measures of follow-on research. For two reasons, this does not appear to be a concern in our empirical setting. First, our point estimates are very similar when we condition on Art Unit-by-year-by-class-by-subclass fixed effects, which suggests that technology specialization based on observables is not a first-order issue in our setting. Second, as we document in Section IVD, we do not uncover evidence of sorting based on other observable proxies for the value of applications.

D. *Investigating Selection*

In order for examiner leniency to be a valid instrumental variable for the likelihood that a given gene patent application is granted a patent, it must satisfy the exclusion restriction: the instrument can only affect follow-on innovation outcomes through the likelihood that a gene is patented. The institutional details described in Section IVA suggest that the assignment of applications to examiners is plausibly random conditional on Art Unit-by-application year fixed effects, lending some a priori credibility to the exclusion restriction. In this section, we empirically assess whether this assumption is reasonable by investigating whether patent applications assigned to "lenient" and "strict" examiners look similar on observable characteristics fixed at the time of patent application.

Ideally, we would empirically assess selection using variables that are correlated with the "patentability" of the application at the time of filing, so that we could test whether applications that appear more patentable tend to be assigned to more lenient examiners. As discussed by Lemley and Sampat (2012), it is difficult to identify variables that measure the "patent-worthiness" of an invention.[39] A variety of metrics have been proposed as measures of the value of granted patents: forward citations (Trajtenberg 1990), patent renewal behavior (Pakes 1986, Schankerman and Pakes 1986, Bessen 2008), patent ownership reassignments (Serrano 2010), patent litigation (Harhoff, Scherer, and Vopel 2003), and excess stock return values (Kogan et al. 2017). For our purposes, these measures are not appropriate: we need measures of patent value that are defined for patent applications (not just for granted patents), and also want a measure that is fixed at the time of application (and hence unaffected by subsequent grant decisions). For these reasons, we focus on two value measures which fit these criteria: patent family size and claims count.[40]

Generally stated, a patent "family" is defined as a set of patent applications filed with different patenting authorities (e.g., United States, Europe, Japan) that refer to the same invention. The key idea is that if there is a per-country cost of filing for a patent, firms will be more likely to file a patent application in multiple countries if they perceive the patent to have higher private value. Past work starting with Putnam (1996) has documented evidence that patent family size is correlated with other measures of patent value. We define patent family size as the number of unique countries in which the patent application was filed.

We use claims count as an alternative value measure that is fixed at the time of patent application, as proposed by Lanjouw and Schankerman (2001). The key idea underlying this measure is that patents list "claims" over specific pieces of intellectual property, and that patents with more claims may be more valuable.

For our purposes, there are two key empirical questions we want to investigate using these measures. First, do patent family size and/or claims count predict patent grant? While clearly imperfect metrics of patentability, these variables are predictive

---

[39] In their paper, Lemley and Sampat (2012) show that two observable characteristics fixed at the time of application—the number of pages in the application and the patent family size—are not correlated with a measure of examiner experience (years of employment at the USPTO). That evidence provides some indirect support for our exclusion restriction.

[40] Unfortunately, backward citations-based measures cannot be constructed for patent applications, because published US patent applications do not list backward citations.

of patent grants: if we regress an indicator variable for patent grant on these two variables, the p-value from an F-test of joint significance is <0.001. Second, is the predicted probability of patent grant—predicted as a function of family size and claims count—correlated with our examiner leniency instrument? If we regress the predicted probability of patent grant (predicted as a function of family size and claims count) on our examiner leniency instrument (residualized by Art Unit-by-application year fixed effects), we estimate a coefficient of 0.013 (standard error 0.003). This relationship is displayed nonparametrically in the plot of lighter yellow triangles in Figure 3: consistent with our economically small regression estimate, there is no visual relationship between the predicted probability of patent grant and our instrument.

Taken together, the analysis in this section provides indirect support of our exclusion restriction in the following sense: we find no evidence that applications which appear more likely to be patented based on covariates that are fixed at the time of filing are differentially assigned to more lenient examiners. Hence, the variation in grant rates across examiners appears to reflect differences in the decisions made on ex ante similar applications.

E. *Instrumental Variables Estimates*

Table 3 documents our instrumental variable estimates, relating patenting (instrumented by examiner leniency) to follow-on innovation outcomes.[41]

For our measures of follow-on scientific research (publications; panel A of Table 3) and commercialization (clinical trials; panel B of Table 3), the 95 percent confidence intervals on our log estimates can reject declines of more than 9 percent; the 95 percent confidence intervals on our binary versions of these outcomes ("any publication" or "any clinical trial") are much less precise. For our measure of diagnostic test availability (panel C of Table 3), the 95 percent confidence interval suggests that, relative to a mean of around 9 percent, we can reject declines in this outcome of greater than 42 percent and reject increases of greater than 11 percent. While some of the individual estimates are statistically significant at conventional levels (for example, the p-value for our estimate of how log scientific publications were affected is 0.024), we emphasize here that the economic magnitudes of these estimates are small.

The estimates from our first quasi-experimental approach—comparing follow-on innovation across genes claimed in successful versus unsuccessful patent applications—are more precise than these estimates. Our confidence in interpreting those estimates as causal is strengthened by the fact that the examiner leniency instrument generates similar results, albeit which are much less precise. From an economic perspective, the estimates from our first approach can cleanly reject the effect sizes documented in the prior literature on how non-patent forms of intellectual property affect follow-on innovation (Williams 2013, Murray et al. 2016). For example, Williams (2013) documents that a non-patent form of intellectual property on the human genome generated declines in follow-on innovation on the order of 30 percent, whereas the estimates from our first approach tend to reject declines or increases in follow-on innovation of more than 5–15 percent. The estimates from

---

[41] Online Appendix Table D.6 documents corresponding ordinary least squares (OLS) estimates for this specification, for comparison.

Table 3—Patents and Follow-On Innovation on Human Genes by Examiner Leniency: Instrumental Variables Estimates

	log of follow-on innovation in 2011–2012	Any follow-on innovation in 2011–2012
	(1)	(2)
Panel A. Scientific publications		
Patent granted (instrumented)	−0.0230	−0.0187
	(0.0102)	(0.0089)
Mean of dependent variable	0.0798	0.0888
Observations	293,652	293,652
Panel B. Clinical trials		
Patent granted (instrumented)	−0.0488	−0.0293
	(0.0209)	(0.0118)
Mean of dependent variable	0.0690	0.0500
Observations	293,652	293,652
Panel C. Diagnostic test		
Patent granted (instrumented)	–	−0.0141
	–	(0.0123)
Mean of dependent variable	–	0.0918
Observations	–	293,652

Notes: This table presents instrumental variable estimates, relating follow-on innovation to whether a patent application was granted a patent, instrumented by our examiner leniency instrument. The sample for these regressions is constructed from application-gene-level data, and includes patent application-gene-level observations in our human gene sample ($N = 293,652$). All regressions include Art Unit-by-application year fixed effects. Each coefficient is from a separate regression. Standard errors are clustered at the patent application level (Inoue and Solon 2010, Pacini and Windmeijer 2016).

our second approach, which are much less precise, can generally but not always reject declines in follow-on innovation on the order of 30 percent.

V. Discussion

A. *Connecting Theory to Data*

A well-developed theoretical literature has generated predictions for how patents will affect follow-on innovation under different conditions. For example, if a patent holder can develop all possible follow-on inventions herself, then all socially desirable follow-on inventions will be developed. However, as stressed by Scotchmer (2004), the patent holder may not know all potential opportunities for follow-on innovation; in Scotchmer's language, ideas for follow-on innovations may be "scarce." The data suggest this is the relevant case in our context: as best we can measure, very little of the follow-on research in our sample is done by the patent assignee.[42] Tracing the identity of follow-on innovators is most feasible for our

---

[42] Note that documenting this type of heterogeneity requires some relatively unusual data construction efforts because published patent applications are not required to list the assignee of the application, which is missing for about 50 percent of published applications. Based on discussions with individuals at the USPTO, we developed an approach which allows us to fill in missing assignee names in most but not all cases; details of this approach are included in online Appendix C.

clinical trials data, which record the name of the organization in charge of the clinical trial.[43] We matched clinical trial organizations with assignee names in the patent application data, and found that among all observed clinical trials in the human gene sample these entities were the same in only 0.1 percent of cases.[44] This suggests the most relevant case in our setting is one in which follow-on innovations require cross-firm licensing agreements.

If ex ante licensing agreements can be signed prior to follow-on inventors sinking their R&D investments, all socially desirable follow-on inventions will still be developed.[45] However, if there are impediments to ex ante licensing, such as asymmetric information, then follow-on innovation will be lower than socially desirable due to the problem of how to divide the profit across firms (Scotchmer 1991, Green and Scotchmer 1995). This profit division problem could be exacerbated if transaction costs hinder cross-firm licensing agreements (Heller and Eisenberg 1998, Anand and Khanna 2000, Bessen 2004). One natural conjecture is that bargaining failures may be less common in cases where one or both parties is a non-profit entity such as a university. We investigated this possibility in our data by manually inspecting each clinical trial organization and patent application assignee in order to classify each as a for-profit or not-for-profit. In essentially all cases (96 percent), patent applications are assigned to for-profit firms.[46] Moreover, in nearly all cases both the assignee and the clinical trial organization are for-profit entities (86 percent of our sample). Unfortunately, this means we lack sufficient variation in our sample to explore heterogeneity in the for-profit status of assignees or follow-on innovators.[47]

Interpreted through the lens of this theoretical literature, the fact that follow-on innovations in this market will often require cross-firm licensing agreements (given the low rates of follow-on innovation by patent holders themselves), together with the fact that we empirically find no evidence that patents induce economically meaningful reductions in follow-on innovation, suggests that licensing contracts in this market operated at least somewhat efficiently.

Importantly, this interpretation depends critically on the assumption that gene patents are sufficiently broad that the types of follow-on inventions we measure would require gene patent licenses. In our view, legal and policy writings on this

---

[43] Tracing the identity of follow-on innovators is much less feasible for our other two outcomes: in the case of scientific publications, publicly available data on scientific publications (in PubMed) only record the institutional affiliation of the first author (when the last would arguably be most relevant), and our diagnostic tests data do not provide information on the institution offering the test.

[44] Specifically, we matched using the Levenshtein distance divided by the average length of the two strings. Pairs with low scores and pairs with at least one word in common were manually checked for false negative matches. Because companies could conceivable operate clinical trials through subsidiaries, we purchased data from ReferenceUSA on the parent companies of each firm in our clinical trials and patent application data in order to classify firms that share the same parent company as belonging to the same "company family." In practice, our results in this section are essentially identical whether or not we account for company families using this data.

[45] Also important to mention is the strand of research dating back at least to Kitch (1977), which has argued, in contrast with this Scotchmer-style literature, that patents may facilitate investment and technology transfers across firms (Arora 1995; Arora, Fosfuri, and Gambardella 2001; Kieff 2001; Gans, Hsu, and Stern 2002, 2008), increasing incentives for follow-on research and commercialization.

[46] Previous research such as Jensen and Murray (2005) documented that public sector assignees accounted for around one-quarter of issued gene patents during an earlier time period than our data cover; this may indicate that the public sector became less prominent in gene patenting over time, though we do not examine this directly.

[47] Galasso and Schankerman (2015) also explore heterogeneity based on the size of the patent assignee. Again, in practice, this dimension of heterogeneity is not meaningful in our sample of human gene patents: 88 percent of the applications in our sample are assigned to one of four firms which would all be reasonably classified as small biotechnology firms over our period of study (Genentech, Corixa, Incyte, and Millennium).

topic clearly support this assumption. In a now-classic paper, Heller and Eisenberg (1998) documented that more than 100 issued US patents included the term "adrenergic receptor" in the claim language, and pointed to this as an example of the complex biomedical patent landscape. Consistent with this concern, in a comment on the Heller-Eisenberg paper published in the same issue of *Science*, the then-USPTO biotechnology examination unit head John Doll argued that gene patents would be interpreted quite broadly by the USPTO (Doll 1998). For example, Doll wrote "… once a [gene] is patented, that patent extends to any use, even those that have not been disclosed in the patent" (p. 690). This interpretation strongly supports the assumption that the follow-on innovations we measure would require gene patent licenses.

However, one can find writings which take the opposite position. For example, in a second comment on the Heller-Eisenberg paper published in the same issue of *Science*, Seide and MacLeod (1998) argued that based on a cursory patent clearance review, at most only a small number of licenses might be required in the adrenergic receptor case. A similar argument has been made in the context of gene patents by Holman (2012): although a large number of human genes are claimed by US patents, he argues that a reading of the claims suggests that not all subsequent research on these genes would necessarily infringe. An anecdote is perhaps helpful in illustrating a concrete example of why gene patent licenses may not always be needed. Andrew Pollack reported in the *New York Times*[48] that pharmaceutical firm Bristol-Myers abandoned research on more than 50 cancer-related proteins due to conflicts with gene patent holders. While interpreted by many as evidence of gene patents deterring follow-on innovation, in fact this example was reported as part of a licensing agreement Bristol-Myers brokered with another firm (Athersys) for a method enabling the use of protein targets without infringing gene patents. That is, in this case, despite Bristol-Myers reporting frictions in licensing markets for gene patents, the follow-on research they wished to pursue was able to occur.

One might think that the question of whether follow-on innovations require licenses could be resolved simply by reading the text of the claims of the gene patents in our sample. However, patent breadth is determined not only by the text of the patent, but also by courts' interpretations of these claims. Because very few gene patents have been litigated in court, we have little data on how broadly these patents would be interpreted in practice.

We want to stress two key take-aways from this discussion of patent breadth. First, from a theoretical perspective narrow patents should not deter follow-on innovation, so to the extent that this alternative interpretation is correct our analysis would be most useful in highlighting that documentation of biomedical patenting frequency (as in Jensen and Murray 2005) should not necessarily be interpreted as evidence of a problem; analyses such as the one in this paper are needed in order to investigate whether patents are sufficiently broad to actually be affecting real economic behavior and outcomes. Second, while these two interpretations are interesting and important to consider from an academic perspective, distinguishing between them is not required from a policy perspective: the fact that gene patents do not appear to

---

[48] Andrew Pollack, "Bristol-Myers and Athersys Make Deal on Gene Patents," *New York Times,* January 8, 2001.

have hindered follow-on innovation is sufficient to inform both the Nordhaus-style theoretical question of interest in this market, and the policy-relevant question at the basis of the recent US Supreme Court *AMP v. Myriad* ruling.

B. *Interpreting Our Estimates: All Intellectual Property Is Not Alike*

Our empirical estimates suggest that gene patents have not had quantitatively important effects on follow-on scientific research nor follow-on product development. This conclusion stands in contrast with a June 2013 US Supreme Court decision which unanimously ruled that human genes should not be patentable because such patents "would 'tie up' … [genes] and… inhibit future innovation" (p. 11). While the Court cited no empirical evidence on this point, and a broad range of commenters (e.g., National Research Council 2006, Caulfield et al. 2013) argued that there was essentially no empirical evidence available to either support or refute that assertion, we here give some context for why the previous literature may have been—perhaps erroneously—interpreted as supporting this view.

Perhaps most closely related to this paper is previous work by one of the authors (Williams 2013), who analyzed a non-patent form of database protection held by the private firm Celera on their version of the sequenced human genome. Celera attempted (but largely failed) to obtain patents on its sequenced genetic data, most likely because Celera's sequenced genes alone were not sufficient to meet the USPTO's utility requirement in the absence of specific knowledge about the functions of those genes in the human body. When they were unable to obtain patents, Celera hired a well-known intellectual property lawyer and asked him to design an alternative form of intellectual property that would be the next best available alternative means of capturing a return on their investment in sequencing the human genome. This lawyer designed a contract-law based form of intellectual property that had several key features. First, Celera's data were "disclosed" in 2001 (Venter et al. 2001), in the sense that any individual could view data on the assembled genome through Celera's website or by obtaining a data DVD from the company. Second, viewing or obtaining Celera's data required agreeing not to redistribute the data. These restrictions on redistribution allowed Celera to grant academic researchers free access to Celera's data for non-commercial research, and to sell its data to larger institutions such as pharmaceutical companies. Although terms of specific deals were private, Service (2001) reports that pharmaceutical companies were paying between $5 million and $15 million a year, whereas universities and nonprofit research organizations were paying between $7,500 and $15,000 for each lab given access to the data. Importantly, these restrictions on redistribution also meant that Celera's data were not disclosed in the main open-access databases (e.g., GenBank) used by researchers over this period. Third, any researcher wanting to use Celera's data for commercial purposes was required to negotiate a licensing agreement with Celera. Williams (2013) documents that this non-patent form of intellectual property on human genes was associated with large declines in follow-on scientific research and commercial product development—on the order of 30 percent—relative to genes that were sequenced at the same time by the publicly-funded Human Genome Project, which placed all of their sequenced genes in public open-access databases such as GenBank.

Why did Celera's intellectual property cause declines in follow-on innovation, while gene patents did not?[49] Theoretical models tend to analyze stylized characterizations of intellectual property rights in which Celera's contract law-based form of intellectual property could reasonably be seen as practically identical to patent protection: both intended to provide an incentive for the discovery of an invention, and both required licensing agreements for any follow-on commercial applications developed by inventors outside of the firm. Seen from this perspective, the contrast in impacts on follow-on innovation between Celera's intellectual property and gene patents is perhaps surprising.

However, viewed through an alternative lens—what one might call the "disclosure" lens—this contrast is less surprising and perhaps even expected. The patent system requires inventors to disclose the claimed invention. Although theoretical models of intellectual property have tended to pay relatively little attention to the disclosure requirement, in our view this is a critical contrast between Celera's intellectual property and gene patents. For gene patent applications, the patent system's disclosure requirement induced explicit reporting of the claimed DNA sequence in a way that enabled open access to the sequence data underlying both accepted and rejected applications for all prospective follow-on users. In contrast, as detailed above Celera's data were disclosed in a much more restrictive way.

Both the institutional details of our context and a growing academic literature suggest that this difference in disclosure may plausibly account for the different observed effects of Celera's intellectual property and gene patents on follow-on innovation outcomes.[50] Over the time period of our study, all DNA sequences claimed in patent applications submitted to the USPTO were disclosed in the main open-access databases (e.g., GenBank) used by researchers.[51] In contrast, the no-redistribution restriction on Celera's sequenced data meant that their data were not added to the key open-access databases that were widely used by researchers. A natural hypothesis is that this lack of open disclosure led to fewer scientific researchers analyzing Celera's data, and that this in turn led to fewer commercializable discoveries. Consistent with this hypothesis, both the theoretical work of Aghion, Dewatripont, and Stein (2008) and other sources of empirical evidence (Furman and Stern 2011, Murray et al. 2016) suggest that limiting access to research materials may discourage follow-on research. Also consistent with this hypothesis, and with our empirical results, Galasso and Schankerman (2015) find no evidence that patent invalidations—which should induce a change in licensing requirements but no change in disclosure—increase patent citations (their measure of follow-on innovation) in the technology field that includes most of our gene patents.[52]

---

[49] Notably, some commenters such as Joseph Stiglitz argued that gene patents were likely to deter follow-on innovation on the basis of the Williams (2013) estimates; see Marshall (2013).

[50] Gallini (2017) makes a similar argument: patents require disclosure in exchange for property rights, whereas non-patent IP does not. She discusses how this intuition can be formalized in a Green and Scotchmer (1995) style model.

[51] As we discuss below, this was not true for rejected patent applications filed before November 29, 2000, but is true for all patent applications in our sample as all were filed after that date.

[52] Specifically, the most common patent class in our sample is 435 (Chemistry: molecular biology and microbiology), which is included in the NBER patent technology field "drugs and medical"; Galasso and Schankerman (2015) do not find evidence that patent invalidations increase follow-on citations in that subcategory.

Our findings provide indirect support for the idea that gene patents effectively disclose information. More precisely, our findings indirectly support the idea that gene patent *applications*—both accepted and rejected, over the time period of our study—effectively disclose information. One limitation of our analysis is that we cannot speak to whether the effect of a gene patent application being accepted versus rejected may have been different for patent applications filed before November 29, 2000, when rejected applications were not required to be publicly disclosed. Under that earlier regime, patent applicants had the option—if their applications were rejected—of pursuing alternative forms of intellectual property protection on their inventions. In the case of Celera, when their patent applications were rejected and *not* required to be disclosed, the firm chose to rely on an alternative form of intellectual property, which, taken at face value, resulted in both lower private returns and lower social returns to their research investment. That is, the lower levels of follow-on innovation on Celera's genes suggest that both Celera and the social planner may have preferred to have granted Celera's gene patent applications rather than rejecting them.

Importantly, this disclosure interpretation reinforces one important reason why our estimates may not generalize beyond the case of human gene patents: the USPTO's specific (and more stringent) requirements for the disclosure of sequenced genetic data may have made the disclosure function particularly effective in our context. While the USPTO requires that all patent applicants satisfy the disclosure requirement (35 USC. §112), it is generally not in patent applicants' interest to disclose details about their invention, as disclosure increases the risk of imitation. For example, Fromer (2009) argues that it is in patentees' interest to disclose as little information as possible in their patent application.[53] Patent attorneys, working on patentees' behalf, arguably have a legal responsibility to minimize disclosure subject to what the USPTO requires. Roin (2005) summarizes the effects of these incentives: "the written descriptions are often drafted in ways that make them tedious or difficult for engineers to understand" (p. 2024), "… the goals of clarity and brevity take a back seat to drafting strategies meant to ensure that patents are interpreted broadly by the courts" (p. 2025). While recent survey evidence paints a somewhat more optimistic view (Ouellette, 2012, 2017),[54] our basic point stands: the USPTO's disclosure requirements for gene patents were much more stringent than is usually the case, providing an important reason why our estimates may not generalize beyond the case of human gene patents.

From a policy perspective, these findings point to a potential unintended consequence of the US Supreme Court's rulings that human genes as well as several other types of technologies—such as business methods and diagnostic tests—are no longer eligible for patent protection. These rulings seem to be based on the idea

---

[53] She writes: "patentees rationally have little or no incentive to offer more information than the patent laws require and have an incentive to obfuscate information they provide whenever possible" (Fromer 2009, pp. 552–53). Cockburn, Kortum, and Stern (2003, footnote 4) argue that university and public sector patents are often written less strategically than are those from private firms.

[54] Ouellette (2012, 2017) document survey evidence suggesting that researchers do rely on patents as a source of technical information. In nanotechnology (Ouellette 2012), 64 percent of researchers surveyed report having read at least part of a patent (other than their own) for a research purpose; in a broader sample of researchers (Ouellette 2017), 79 percent of industry researchers and 80 percent of researchers in other sectors report having read patents for some scientific reasons, including looking for both general information and specific solutions.

that if these inventions are not patented, they will be placed in the public domain. However, the alternative to patent protection may instead be non-patent intellectual property that could generate worse outcomes than allowing patent protection. As a concrete example, in response to the *Mayo v. Prometheus* ruling which rendered many medical diagnostic methods unpatentable, the law firm Goodwin Proctor issued an "alert" arguing that the ruling called for "a renewed look at trade secret law for protection of diagnostics innovations as an alternative to patents."[55] This type of behavioral response would be critical to assessing the potential welfare implications of the US Supreme Court's decisions on many types of technologies which have recently been declared to be unpatentable.

VI. Conclusion

The contribution of this paper is to investigate whether patents on human genes have affected follow-on scientific research and product development. Using administrative data on successful and unsuccessful patent applications submitted to the US Patent and Trademark Office (USPTO), we link the exact gene sequences claimed in each patent application with data measuring gene-related scientific research (publications) and commercial investments (clinical development). Building on this data, we develop two new sources of quasi-experimental variation: first, a simple comparison of follow-on innovation across genes claimed in successful versus unsuccessful patent applications; and second, use of the "leniency" of the assigned patent examiner as an instrumental variable for whether the patent application was granted a patent. Both approaches suggest that—on average—gene patents have not had quantitatively important effects on follow-on innovation.

This empirical evidence speaks against two existing views. First, there has been widespread concern that patents on human genes may hinder follow-on innovation. For example, in the 2013 *Association for Molecular Pathology v. Myriad Genetics* case, the US Supreme Court invalidated patent claims on genomic DNA, arguing that such patents "would 'tie up' the use of such tools and thereby inhibit future innovation premised upon them" (p. 11). Our empirical estimates do not provide support for patents inducing economically meaningful reductions in follow-on innovation in the context of human genes. Second, dating back at least to the academic work of Kitch (1977), many have argued that patents on basic discoveries play an important role in facilitating subsequent investment and commercialization. Our empirical estimates do not provide support for patents spurring follow-on innovation in the context of human genes.

Taken together with the prior literature, our evidence suggests two conclusions. First, for the case of human genes, the traditional patent trade-off of ex ante incentives versus deadweight loss may be sufficient to analyze optimal patent policy design, because any effects of patents on follow-on innovation appear to be quantitatively small. Second, our evidence, together with the evidence from Williams (2013) on how a non-patent form of intellectual property on the human genome affected follow-on innovation, suggests a somewhat nuanced conclusion: while patent

---

[55] See http://www.goodwinlaw.com/publications/2012/03/lessons-from-mayo-v-prometheus-assessing-patent-ability-and-obtaining-patent-protection.

protection on human genes does not appear to have hindered follow-on innovation, an alternative non-patent form of intellectual property—which was used by a private firm after its gene patent applications were largely unsuccessful in obtaining patent grants—induced substantial declines in follow-on scientific research and product development. This pattern of evidence suggests that changes to patent policy must carefully consider what strategies that firms will use to protect their discoveries in the absence of patents, and that an understanding of the relative costs and benefits of patent protection compared to those outside options is needed in order to evaluate the welfare effects of patent policy changes.

REFERENCES

Acemoglu, Daron, and Joshua Linn. 2004. "Market Size in Innovation: Theory and Evidence from the Pharmaceutical Industry." *Quarterly Journal of Economics* 119 (3): 1049–90.

Aghion, Philippe, Mathias Dewatripont, and Jeremy C. Stein. 2008. "Academic Freedom, Private-Sector Focus, and the Process of Innovation." *RAND Journal of Economics* 39 (3): 617–35.

Aizer, Anna, Shari Eli, Joseph Ferrie, and Adriana Lleras-Muney. 2016. "The Long-Run Impact of Cash Transfers to Poor Families." *American Economic Review* 106 (4): 935–71.

Anand, Bharat N., and Tarun Khanna. 2000. "The Structure of Licensing Contracts." *Journal of Industrial Economics* 48 (1): 103–35.

Angrist, Joshua D., Guido W. Imbens, and Alan B. Krueger. 1999. "Jackknife Instrumental Variables Estimation." *Journal of Applied Econometrics* 14 (1): 57–67.

Angrist, Joshua D., and Alan B. Krueger. 1992. "The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables with Moments from Two Samples." *Journal of the American Statistical Association* 87 (418): 328–36.

Arora, Ashish. 1995. "Licensing Tacit Knowledge: Intellectual Property Rights and the Market for Know-How." *Economics of Innovation and New Technology* 4 (1): 41–59.

Arora, Ashish, Andrea Fosfuri, and Alfonso Gambardella. 2001. *Markets for Technology: The Economics of Innovation and Corporate Strategy*. Cambridge, MA: MIT Press.

Bacon, Neil, Doug Ashton, Richard Jefferson, and Marie Connett. 2006. "Biological Sequences Named and Claimed in US Patents and Patent Applications: CAMBIA Patent Lens OS4 Initiative." http://www.cambia.org/daisy/patentlens/2906.html (accessed January 2, 2012).

Berman, Richard, and Amy Schoenhard. 2004. "The Level of Disclosure Necessary for Patent Protection of Genetic Innovations." *Nature Biotechnology* 22: 1307–08.

Bessen, James. 2004. "Holdup and Licensing of Cumulative Innovations with Private Information." *Economics Letters* 82 (3): 321–26.

Bessen, James. 2008. "The Value of U.S. Patents by Owner and Patent Characteristics." *Research Policy* 37 (5): 932–45.

Bound, John. 1989. "The Health and Earnings of Rejected Disability Insurance Applicants." *American Economic Review* 79 (3): 482–503.

Carley, Michael, Deepak Hegde, and Alan Marco. 2015. "What Is the Probability of Receiving a U.S. Patent?" *Yale Journal of Law and Technology* 17 (1): 203–23.

Caulfield, Timothy, Subhashini Chandrasekharan, Yann Joly, and Robert Cook-Deegan. 2013. "Harm, Hype and Evidence: ELSI Research and Policy Guidance." *Genome Medicine* 5 (21).

Cockburn, Iain, Samuel Kortum, and Scott Stern. 2003. "Are All Patent Examiners Equal? Examiners, Patent Characteristics, and Litigation Outcomes." In *Patents in the Knowledge-Based Economy*, edited by Wesley M. Cohen and Stephen A. Merrill, 19–53. Washington, DC: National Academies Press.

Davidson, Russell, and James G. MacKinnon. 2006. "The Case against JIVE." *Journal of Applied Econometrics* 21 (6): 827–33.

Doll, John J. 1998. "The Patenting of DNA." *Science* 280 (5364): 689–90.

Doyle, Joseph J., Jr. 2007. "Child Protection and Child Outcomes: Measuring the Effects of Foster Care." *American Economic Review* 97 (5): 1583–610.

Doyle, Joseph J., Jr. 2008. "Child Protection and Adult Crime: Using Investigator Assignment to Estimate Causal Effects of Foster Care." *Journal of Political Economy* 116 (4): 746–70.

Farre-Mensa, Joan, Deepak Hegde, and Alexander Ljungqvist. 2017. "What Is a Patent Worth? Evidence from the U.S. Patent 'Lottery.'" NBER Working Paper 23268.

Feng, Josh, and Xavier Jaravel. 2018. "Crafting Intellectual Property Rights: Implications for Patent Assertion Entities, Litigation, and Innovation." Unpublished.

Finkelstein, Amy. 2004. "Static and Dynamic Effects of Health Policy: Evidence from the Vaccine Industry." *Quarterly Journal of Economics* 119 (2): 527–64.

Frakes, Michael D., and Melissa F. Wasserman. 2017. "Is the Time Allocated to Review Patent Applications Inducing Examiners to Grant Invalid Patents? Evidence from Microlevel Application Data." *Review of Economics and Statistics* 99 (3): 550–63.

Fromer, Jeanne C. 2009. "Patent Disclosure." *Iowa Law Review* 94 (2): 539–606.

Furman, Jeffrey L., and Scott Stern. 2011. "Climbing Atop the Shoulders of Giants: The Impact of Institutions on Cumulative Research." *American Economic Review* 101 (5): 1933–63.

Galasso, Alberto, and Mark Schankerman. 2015. "Patents and Cumulative Innovation: Causal Evidence from the Courts." *Quarterly Journal of Economics* 130 (1): 317–69.

Gallini, Nancy. 2017. "Do Patents Work? Thickets, Trolls, and Antibiotic Resistance." *Canadian Journal of Economics* 50 (4): 893–926.

Gans, Joshua S., David H. Hsu, and Scott Stern. 2002. "When Does Start-up Innovation Spur the Gale of Creative Destruction?" *RAND Journal of Economics* 33 (4): 571–86.

Gans, Joshua S., David H. Hsu, and Scott Stern. 2008. "The Impact of Uncertain Intellectual Property Rights on the Market for Ideas: Evidence from Patent Grant Delays." *Management Science* 54 (5): 982–97.

Gaule, Patrick. 2015. "Patents and the Success of Venture-Capital Backed Startups: Using Examiner Assignment to Estimate Causal Effects." Unpublished.

Graham, Stuart J. H., and Deepak Hegde. 2013. "Do Inventors Value Secrecy in Patenting? Evidence from the American Inventor's Protection Act of 1999." Unpublished.

Green, Jerry R., and Suzanne Scotchmer. 1995. "On the Division of Profit in Sequential Innovation." *RAND Journal of Economics* 26 (1): 20–33.

Greene, William. 2001. "Estimating Econometric Models with Fixed Effects." Unpublished.

Harhoff, Dietmar, Frederic M. Scherer, and Katrin Vopel. 2003. "Citations, Family Size, Opposition and the Value of Patent Rights." *Research Policy* 32 (8): 1343–63.

Heckman, James. 1981. "The Incidental Parameters Problem and the Problem of Initial Condition in Estimating a Discrete Time-Discrete Data Stochastic Process." In *Structural Analysis of Discrete Data with Econometric Applications,* edited by Charles F. Manski and Daniel L. McFadden. Cambridge, MA: MIT Press.

Heller, Michael A., and Rebecca S. Eisenberg. 1998. "Can Patents Deter Innovation? The Anticommons in Biomedical Research." *Science* 280 (5364): 698–701.

Holman, Christopher M. 2012. "Debunking the Myth that Whole-Genome Sequencing Infringes Thousands of Gene Patents." *Nature Biotechnology* 30: 240–44.

Inoue, Atsushi, and Gary Solon. 2010. "Two-Sample Instrumental Variables Estimators." *Review of Economics and Statistics* 92 (3): 557–61.

Jensen, Kyle, and Fiona Murray. 2005. "Intellectual Property Landscape of the Human Genome." *Science* 310 (5746): 239–40.

Kesselheim, Aaron S., Robert M. Cook-Deegan, David Winickoff, and Michelle M. Mello. 2013. "Gene Patenting: The Supreme Court Finally Speaks." *New England Journal of Medicine* 369 (9): 869–75.

Kieff, F. Scott. 2001. "Property Rights and Property Rules for Commercializing Inventions." *Minnesota Law Review* 85: 697–754.

Kitch, Edmund W. 1977. "The Nature and Function of the Patent System." *Journal of Law and Economics* 20 (2): 265–90.

Kline, Patrick, Neviana Petkova, Heidi Williams, and Owen Zidar. 2017. "Who Profits from Patents? Rent-Sharing at Innovative Firms." Unpublished.

Kling, Jeffrey R. 2006. "Incarceration Length, Employment, and Earnings." *American Economic Review* 96 (3): 863–76.

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman. 2017. "Technological Innovation, Resource Allocation, and Growth." *Quarterly Journal of Economics* 132 (2): 665–712.

Lanjouw, Jean O., and Mark Schankerman. 2001. "Characteristics of Patent Litigation: A Window on Competition." *RAND Journal of Economics* 32 (1): 129–51.

Lee, Byungwook, Taehyung Kim, Seon-Kyu Kim, Kwang H. Lee, and Doheon Lee. 2007. "Patome: A Database Server for Biological Sequence Annotation and Analysis in Issued Patents and Published Patent Applications." *Nucleic Acids Research* 35: D47–50.

Lemley, Mark, and Bhaven Sampat. 2008. "Is the Patent Office a Rubber Stamp?" *Emory Law Journal* 58: 181–203.

Lemley, Mark, and Bhaven Sampat. 2010. "Examining Patent Examination." *Stanford Technology Law Review* 2.

Lemley, Mark A., and Bhaven Sampat. 2012. "Examiner Characteristics and Patent Office Outcomes." *Review of Economics and Statistics* 94 (3): 817–27.

Lichtman, Douglas. 2004. "Rethinking Prosecution History Estoppel." *University of Chicago Law Review* 71 (1): 151–82.

Maestas, Nicole, Kathleen J. Mullen, and Alexander Strand. 2013. "Does Disability Insurance Receipt Discourage Work? Using Examiner Assignment to Estimate Causal Effects of SSDI Receipt." *American Economic Review* 103 (5): 1797–829.

Marshall, Eliot. 2013. "Lock Up the Genome, Lock Down Research." *Science* 342 (6154): 72–73.

Merges, Robert P. 1988. "Commercial Success and Patent Standards: Economic Perspectives on Innovation." *California Law Review* 76 (4): 803–76.

Merges, Robert P. 1999. "As Many as Six Impossible Patents before Breakfast: Property Rights for Business Concepts and Patent System Reform." *Berkeley Technology Law Journal* 14: 577–615.

Moon, Seongwuk. 2011. "How Does the Management of Research Impact the Disclosure of Knowledge? Evidence from Scientific Publications and Patenting Behavior." *Economics of Innovation and New Technology* 20 (1–2): 1–32.

Moser, Petra. 2012. "Innovation without Patents: Evidence from World's Fairs." *Journal of Law and Economics* 55 (1): 43–74.

Murray, Fiona, Philippe Aghion, Mathias Dewatripont, Julian Kolev, and Scott Stern. 2016. "Of Mice and Academics: Examining the Effect of Openness on Innovation." *American Economic Journal: Economic Policy* 8 (1): 212–52.

National Research Council. 2006. *Reaping the Benefits of Genomic and Proteomic Research: Intellectual Property Rights, Innovation, and Public Health.* Washington, DC: National Academies Press.

Nordhaus, William. 1969. *Invention, Growth, and Welfare: A Theoretical Treatment of Technological Change.* Cambridge, MA: MIT Press.

Ouellette, Lisa Larrimore. 2012. "Do Patents Disclose Useful Information?" *Harvard Journal of Law and Technology* 25 (2): 546–607.

Ouellette, Lisa Larrimore. 2017. "Who Reads Patents?" *Nature Biotechnology* 35: 421–24.

Pacini, David, and Frank Windmeijer. 2016. "Robust Inference for the Two-Sample 2SLS Estimator." *Economics Letters* 146: 50–54.

Pakes, Ariel S. 1986. "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks." *Econometrica* 54 (4): 755–84.

Putnam, Jonathan. 1996. "The Value of International Patent Rights." PhD diss. Yale University.

Rai, Arti K. 2012. "Patent Validity across the Executive Branch: Ex Ante Foundations for Policy Development." *Duke Law Journal* 61: 1237–81.

Righi, Cesare, and Timothy Simcoe. 2017. "Patent Examiner Specialization." NBER Working Paper 23913.

Roin, Benjamin. 2005. "The Disclosure Function of the Patent System (or Lack Thereof)." *Harvard Law Review* 118 (6): 2007–28.

Sampat, Bhaven, and Heidi L. Williams. 2019. "How Do Patents Affect Follow-On Innovation? Evidence from the Human Genome: Dataset." *American Economic Review.* https://doi.org/10.1257/aer.20151398.

Schankerman, Mark, and Ariel Pakes. 1986. "Estimates of the Value of Patent Rights in European Countries During the Post-1950 Period." *Economic Journal* 96 (384): 1052–76.

Scherer, Frederic. 2002. "The Economics of Human Gene Patents." *Academic Medicine* 77 (12): 1348–67.

Scotchmer, Suzanne. 1991. "Standing on the Shoulders of Giants: Cumulative Research and the Patent Law." *Journal of Economic Perspectives* 5 (1): 29–41.

Scotchmer, Suzanne. 2004. *Innovation and Incentives.* Cambridge, MA: MIT Press.

Seide, Rochelle, and Janet MacLeod. 1998. "Comment on Heller and Eisenberg." *ScienceOnline.* https://web.archive.org/web/20041215091350/http://www.sciencemag.org:80/feature/data/980465/seide.shl (accessed November 21, 2018).

Serrano, Carlos J. 2010. "The Dynamics of the Transfer and Renewal of Patents." *RAND Journal of Economics* 41 (4): 686–708.

Service, Robert F. 2001. "Can Data Banks Tally Profits?" *Science* 291 (5507): 1203.

Stock, James H., Jonathan H. Wright, and Motohiro Yogo. 2002. "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments." *Journal of Business and Economic Statistics* 20 (4): 518–29.

Trajtenberg, Manuel. 1990. *Economic Analysis of Product Innovation: The Case of CT Scanners.* Harvard Economic Studies, Vol. 160. Cambridge, MA: Harvard University Press.

United States Government Accountability Office. 2005. *Intellectual Property: USPTO Has Made Progress in Hiring Examiners, but Challenges to Retention Remain.* Washington, DC: US Government Accountability Office.

United States Patent and Trademark Office. 2001. "Utility Examination Guidelines." *Federal Register* 66 (4): 1092–99.

United States Supreme Court. 2013. *Association for Molecular Pathology et al. v. Myriad Genetics Inc. et al.* 12-398.

Venter, J. Craig, Mark D. Adams, Eugene W. Myers, Peter W. Li, Richard J. Mural, Granger G. Sutton, and Hamilton O. Smith. 2001. "The Sequence of the Human Genome." *Science* 291 (5507): 1304–51.

von Wachter, Till, Jae Song, and Joyce Manchester. 2011. "Trends in Employment and Earnings of Allowed and Rejected Applicants to the Social Security Disability Insurance Program." *American Economic Review* 101 (7): 3308–29.

Williams, Heidi L. 2013. "Intellectual Property Rights and Innovation: Evidence from the Human Genome." *Journal of Political Economy* 121 (1): 1–27.

Wolinsky, Scott. 2002. "An Inside Look at the Patent Examination Process." *The Metropolitan Corporate Counsel* 10 (9): 18.

This article has been cited by:

1. Benjamin Büttner, Murat Firat, Emilio Raiteri. 2022. Patents and knowledge diffusion. *Research Policy* **51**:10, 104584. [Crossref]

2. Inchae Park, Giorgio Triulzi, Christopher L. Magee. 2022. Tracing the emergence of new technology: A comparative analysis of five technological domains. *Technological Forecasting and Social Change* **184**, 122014. [Crossref]

3. Laurent R Bergé, Thorsten Doherr, Katrin Hussinger. 2022. How patent rights affect university science. *Industrial and Corporate Change* **98**. . [Crossref]

4. Anna Mikusheva, Liyang Sun. 2022. Inference with Many Weak Instruments. *The Review of Economic Studies* **89**:5, 2663-2686. [Crossref]

5. Benjamin Barber, Luis Diestre. 2022. Can firms avoid tough patent examiners through examiner-shopping? Strategic timing of citations in USPTO patent applications. *Strategic Management Journal* **43**:9, 1854-1871. [Crossref]

6. Lucy Xiaolu Wang. 2022. Global drug diffusion and innovation with the medicines patent pool. *Journal of Health Economics* **85**, 102671. [Crossref]

7. Christian Fisch, Michele Meoli, Silvio Vismara, Jörn H. Block. 2022. The effect of trademark breadth on IPO valuation and post-IPO performance: an empirical investigation of 1510 European IPOs. *Journal of Business Venturing* **37**:5, 106237. [Crossref]

8. Margaret K. Kyle. 2022. Incentives for pharmaceutical innovation: What's working, what's lacking. *International Journal of Industrial Organization* **84**, 102850. [Crossref]

9. Walker Hanlon, Taylor Jaworski. 2022. Spillover Effects of Intellectual Property Protection in the Interwar Aircraft Industry. *The Economic Journal* **132**:645, 1824-1851. [Crossref]

10. Daniel P. Gross. 2022. The Hidden Costs of Securing Innovation: The Manifold Impacts of Compulsory Invention Secrecy. *Management Science* **23**. . [Crossref]

11. Deepak Hegde, Alexander Ljungqvist, Manav Raj. 2022. Quick or Broad Patents? Evidence from U.S. Startups. *The Review of Financial Studies* **35**:6, 2705-2742. [Crossref]

12. Haoyang Song, Jianhua Hou, Yang Zhang. 2022. Patent protection: does it promote or inhibit the patented technological knowledge diffusion?. *Scientometrics* **127**:5, 2351-2379. [Crossref]

13. Elisabeth Kempf, Oliver Spalt. 2022. Attracting the Sharks: Corporate Innovation and Securities Class Action Lawsuits. *Management Science* **83**. . [Crossref]

14. Elliott Ash, Daniel L. Chen, Sergio Galletta. 2022. Measuring Judicial Sentiment: Methods and Application to US Circuit Courts. *Economica* **89**:354, 362-376. [Crossref]

15. Jinyoung Kim. 2022. Teamwork in innovation under time pressure. *Labour Economics* **75**, 102137. [Crossref]

16. Anthony Bald, Eric Chyn, Justine Hastings, Margarita Machelett. 2022. The Causal Impact of Removing Children from Abusive and Neglectful Homes. *Journal of Political Economy* . [Crossref]

17. Bin Yang, Tao Yuan. 2022. Trademark and IPO underpricing. *Financial Management* **51**:1, 271-296. [Crossref]

18. Maher Kooli, Aoran Zhang, Yunfei Zhao. 2022. How IPO firms' product innovation strategy affects the likelihood of post-IPO acquisitions?. *Journal of Corporate Finance* **72**, 102159. [Crossref]

19. Shiri Mermelstein, Hilde Stevens. 2022. TRIPS to Where? A Narrative Review of the Empirical Literature on Intellectual Property Licensing Models to Promote Global Diffusion of Essential Medicines. *Pharmaceutics* **14**:1, 48. [Crossref]

20. Kevin A. Bryan, Yasin Ozcan. 2021. The Impact of Open Access Mandates on Invention. *The Review of Economics and Statistics* **103**:5, 954-967. [Crossref]

21. Charles A.W. deGrazia, Nicholas A. Pairolero, Mike H.M. Teodorescu. 2021. Examination incentives, learning, and patent office outcomes: The use of examiner's amendments at the USPTO. *Research Policy* **50**:10, 104360. [Crossref]

22. Barbara Biasi, Petra Moser. 2021. Effects of Copyrights on Science: Evidence from the WWII Book Republication Program. *American Economic Journal: Microeconomics* **13**:4, 218-260. [Abstract] [View PDF article] [PDF with links]

23. Seokbeom Kwon, Alan C. Marco. 2021. Can antitrust law enforcement spur innovation? Antitrust regulation of patent consolidation and its impact on follow-on innovations. *Research Policy* **50**:9, 104295. [Crossref]

24. Guoguo Zhang, Honghong Zhang. 2021. Human mobility restrictions and innovation: evidence from China. *Asian Journal of Technology Innovation* **55**, 1-26. [Crossref]

25. John Liechty, Stefan Wuyts. 2021. 'If I had a hedge fund, I would cure diabetes': endogenous mechanisms for creating public goods. *SN Business & Economics* **1**:10. . [Crossref]

26. Michael D. Frakes, Melissa F. Wasserman. 2021. Knowledge spillovers, peer effects, and telecommuting: Evidence from the U.S. Patent Office. *Journal of Public Economics* **198**, 104425. [Crossref]

27. Jeffrey M. Kuhn, Mike H. M. Teodorescu. 2021. The Track One Pilot Program: Who benefits from prioritized patent examination?. *Strategic Entrepreneurship Journal* **15**:2, 185-208. [Crossref]

28. Michael J. Andrews. 2021. Historical patent data: A practitioner's guide. *Journal of Economics & Management Strategy* **30**:2, 368-397. [Crossref]

29. Colleen Honigsberg, Matthew Jacob. 2021. Deleting misconduct: The expungement of BrokerCheck records. *Journal of Financial Economics* **139**:3, 800-831. [Crossref]

30. Vinícius Eduardo Ferrari, José Maria Ferreira Jardim da Silveira, Maria Ester Soares Dal-Poz. 2021. Patent network analysis in agriculture: a case study of the development and protection of biotechnologies. *Economics of Innovation and New Technology* **30**:2, 111-133. [Crossref]

31. Matthew T. Gustafson, Ivan T. Ivanov, Ralf R. Meisenzahl. 2021. Bank monitoring: Evidence from syndicated loans. *Journal of Financial Economics* **139**:2, 452-477. [Crossref]

32. Deepak Hegde, Alexander Ljungqvist, Manav Raj. 2021. Race, Glass Ceilings, and Lower Pay for Equal Work. *SSRN Electronic Journal* **119**. . [Crossref]

33. Steffen Juranek, Håkon Otneim. 2021. Using machine learning to predict patent lawsuits. *SSRN Electronic Journal* **92**. . [Crossref]

34. Kevin A. Bryan, Heidi L. Williams. Innovation: market failures and public policies 281-388. [Crossref]

35. Pu-yan Nie, Zi-rui Chen, Chan Wang. 2021. Intellectual property pricing under asymmetric duopoly. *Journal of Retailing and Consumer Services* **58**, 102261. [Crossref]

36. Eduardo Melero, Neus Palomeras, David Wehrheim. 2020. The Effect of Patent Protection on Inventor Mobility. *Management Science* **66**:12, 5485-5504. [Crossref]

37. Martin Watzinger, Thomas A. Fackler, Markus Nagler, Monika Schnitzer. 2020. How Antitrust Enforcement Can Spur Innovation: Bell Labs and the 1956 Consent Decree. *American Economic Journal: Economic Policy* **12**:4, 328-359. [Abstract] [View PDF article] [PDF with links]

38. Yarden Katz, Ulrich Matter. 2020. Metrics of Inequality: The Concentration of Resources in the U.S. Biomedical Elite. *Science as Culture* **29**:4, 475-502. [Crossref]

39. Corinne Langinier, Stéphanie Lluis. 2020. DEPARTURE AND PROMOTION OF U.S. PATENT EXAMINERS: DO PATENT CHARACTERISTICS MATTER?. *Contemporary Economic Policy* **95**. . [Crossref]

40. Dylan Roskams-Edris. 2020. Intellectual Property Policy at The Neuro, an Open Science Institute. *Qeios* . [Crossref]

41. Tedi Skiti. 2020. Institutional entry barriers and spatial technology diffusion: Evidence from the broadband industry. *Strategic Management Journal* **41**:7, 1336-1361. [Crossref]

42. Daniel E. Ho, Lisa Larrimore Ouellette. 2020. Improving Scientific Judgments in Law and Government: A Field Experiment of Patent Peer Review. *Journal of Empirical Legal Studies* **17**:2, 190-223. [Crossref]

43. Sonja Lück, Benjamin Balsmeier, Florian Seliger, Lee Fleming. 2020. Early Disclosure of Invention and Reduced Duplication: An Empirical Test. *Management Science* **66**:6, 2677-2685. [Crossref]

44. Ryan Hill, Carolyn Stein, Heidi Williams. 2020. Internalizing Externalities: Designing Effective Data Policies. *AEA Papers and Proceedings* **110**, 49-54. [Abstract] [View PDF article] [PDF with links]

45. Markus Nagler, Stefan Sorg. 2020. The disciplinary effect of post-grant review – Causal evidence from European patent opposition. *Research Policy* **49**:3, 103915. [Crossref]

46. JOAN FARRE-MENSA, DEEPAK HEGDE, ALEXANDER LJUNGQVIST. 2020. What Is a Patent Worth? Evidence from the U.S. Patent "Lottery". *The Journal of Finance* **75**:2, 639-682. [Crossref]

47. David Argente, Salome Baslandze, Douglas Hanley, Sara Moreira. 2020. Patents to Products: Product Innovation and Firm Dynamics. *SSRN Electronic Journal* . [Crossref]

48. Liu Hao. 2020. Patent Citation Network Analysis. *SSRN Electronic Journal* . [Crossref]

49. Joseph Raffiee, Florenta Teodoridis. 2020. Does the Political Ideology of Patent Examiners Matter? An Empirical Investigation. *SSRN Electronic Journal* . [Crossref]

50. Cesare Righi, Timothy S. Simcoe. 2020. Patenting Inventions or Inventing Patents? Strategic Use of Continuations at The USPTO. *SSRN Electronic Journal* . [Crossref]

51. Ce Matthew Shi, Yucong Zhao. 2020. Research Exemption and Pharmaceutical Innovation: Evidence from China. *SSRN Electronic Journal* . [Crossref]

52. Charu Gupta. 2020. One product, many patents: Imperfect intellectual property rights in the pharmaceutical industry. *SSRN Electronic Journal* **25**. . [Crossref]

53. Emerson Escolar, Yasuaki Hiraoka, Mitsuru Igami, Yasin Ozcan. 2020. Mapping Firms' Locations in Technological Space: A Topological Analysis of Patent Statistics. *SSRN Electronic Journal* **31**. . [Crossref]

54. Travis Dyer, Stephen Glaeser, Mark H. Lang, Caroline Sprecher. 2020. The Effect of Patent Disclosure Quality on Innovation. *SSRN Electronic Journal* . [Crossref]

55. Nicholas Bloom, John Van Reenen, Heidi Williams. 2019. A toolkit of policies to promote innovation. *Voprosy Ekonomiki* :10, 5-31. [Crossref]

56. Nicholas Bloom, John Van Reenen, Heidi Williams. 2019. A Toolkit of Policies to Promote Innovation. *Journal of Economic Perspectives* **33**:3, 163-184. [Abstract] [View PDF article] [PDF with links]

57. Julian Boulanger. 2019. The Examination of Continuation Applications and the Problem of Invalid Patents in the U.S. *SSRN Electronic Journal* . [Crossref]

58. Julian Boulanger. 2019. Patents and Follow-On Innovation: Evidence From Patent Renewal Decisions. *SSRN Electronic Journal* . [Crossref]

59. Gaétan de Rassenfosse, Kyle Higham. 2019. Decentralising the Patent System. *SSRN Electronic Journal* . [Crossref]

60. Deepak Hegde, Manav Raj. 2019. Does Gender Affect Work? Evidence from U.S. Patent Examination. *SSRN Electronic Journal* . [Crossref]

61. Pranav Desai. 2019. Biased Regulators: Evidence from Patent Examiners. *SSRN Electronic Journal* . [Crossref]

62. Michael Andrews. 2019. Comparing Historical Patent Datasets. *SSRN Electronic Journal* . [Crossref]

63. Aditya Karanam, Deepa Mani, Rajib Lochan Saha. 2019. Ties that Bind: A Network Approach to Assessing Technology Spillovers from the ICT Industry and Their Impact on R&D. *SSRN Electronic Journal* . [Crossref]

64. Sergio Galletta, Elliott Ash, Daniel L. Chen. 2019. Do Judicial Sentiments Affect Social Attitudes?. *SSRN Electronic Journal* . [Crossref]

65. Thomas J. Chemmanur, Manish Gupta, Karen Simonyan. 2018. How Does Private Firm Innovation Affect Anti-Takeover Provisions in Corporate Charters? Evidence from Firms Going Public. *SSRN Electronic Journal* . [Crossref]

66. Gaétan de Rassenfosse, Paul H. Jensen, T'Mir Julius, Alfons Palangkaraya, Elizabeth M. Webster. 2018. Getting Patents: Does the Quality of Patent Attorney Matter?. *SSRN Electronic Journal* **67**. . [Crossref]

67. Eduardo Melero, Neus Palomeras, David Wehrheim. 2017. The Effect of Patent Protection on Inventor Mobility. *SSRN Electronic Journal* . [Crossref]