University Licensing and the Flow of Scientific Knowledge

Neil C. Thompson
Sloan School of Management
Massachusetts Institute of Technology

Arvids A. Ziedonis
Questrom School of Business
Boston University
and
Harvard Business School—Visiting Scholar

David C. Mowery
Haas School of Business
University of California, Berkeley

February 8, 2016

Abstract

As university involvement in technology transfer and entrepreneurship has increased, concerns over the patenting and licensing of scientific discoveries have grown. This paper examines the effect that the licensing of academic patents has on journal citations to academic publications covering the same scientific research. We analyze data on invention disclosures, patents, and licenses from the University of California, a leading U.S. academic patenter and licensor, between 1997 and 2007. We also develop a novel "inventor–based" maximum–likelihood matching technique to automate and generalize Murray's (2002) "patent–paper pairs" methodology. We use this methodology to identify the scientific publications associated with University of California patents and licenses.

Based on a "differences–in–differences" analysis, we find that, in general, licenses are associated with an increase in journal citations to related scientific publications. The timing of this effect supports recent research that suggests that academic licenses may act as positive signals of research potential in research fields linked to the licensed invention (Drivas et al. 2014). In contrast, we find that licensing of research inputs (which we identify through the use of material transfer agreements, or MTAs) depresses citations to related scientific publications.

Our results suggest that, overall, licensing of academic patents does not limit scientific communication linked to patented academic research. But our findings on the effects of licenses on research inputs, however, raise the possibility that licensing may restrict the flow of inputs to further scientific research among researchers.

We are grateful to Wendy Streitz, Ron Franck, and other employees of the UC Office of the President for access to the intellectual property data employed in this paper, and to Thompson Reuters for their data on scientific publications. We also thank Fiona Murray and Scott Stern for sharing their data. This work has greatly benefited from comments and suggestions by Janet Bercovitz, Paul David, Michael Frakes, Bronwyn Hall, Jennifer Kuan, Fiona Murray, Joanne Oxley, Rob Seamans, Heidi Williams, Brian Wright, Rosemarie Ziedonis, participants in seminars at the Copenhagen Business School, UC Berkeley, the Ohio State University, Santa Clara University, the University of South Carolina, Stanford University, SUNY Buffalo, the 2014 USPTO–Searle Center Conference on Innovation Economics at Northwestern University, 2011 NBER Economics of Intellectual Property Policy and Innovation Workshop, the 2011 Georgia Tech Roundtable for Engineering Entrepreneurship Research, and the 2012 Academy of Management annual meeting, and the 2014 Economics of Knowledge Contribution and Distribution Workshop. Robert Picard provided expert programming assistance. The authors gratefully acknowledge financial support from the Institute for Business Innovation at the Haas School of Business, the National Science Foundation, the Industrial Partnerships Office at Lawrence Livermore National Laboratory, the Program in Law, Economics, and Technology at the University of Michigan Law School, the Rotman School of Management at the University of Toronto, and the Sloan Foundation. We also appreciate the support of the Stanford Institute of Economic Policy Research.

1 Introduction

Growth in patenting of academic research advances in U.S. and other universities within the OECD has triggered considerable debate since at least 1980, the year of passage of the U.S. Bayh–Dole Act. Supporters contend that patenting and licensing of university inventions can speed the transfer of scientific discoveries to the private sector, promoting the commercialization of such advances. On the other hand, critics predict a collision between the norms of science and the norms of commerce, arguing that the exclusionary effects of patents will slow the progress of science. Despite the importance of this question, there has been relatively little empirical work on the extent to which the patenting of academic research results affects scientific research progress. The modest volume of such work (e.g., Fehder et al. (2014), Murray and Stern (2007)) reflects the challenge of obtaining data that can be used to examine such effects.

This paper develops a new approach to matching scientific publications and patents, and employs this methodology to examine the effect of licensing on the journal citations to related publications. We interpret increased citations to these scientific publications after a license issues as evidence that licensing is correlated with a positive effect on the prominence and use by other scientists of the knowledge embodied in the paper, while decreases in such citations may indicate a restrictive effect of licensing on scientific communication. Drawing on related work by Walsh et al. (2007), we also examine the effect of licenses on related scientific publications covering the inputs to the experiments of other researchers, a class of knowledge often referred to as "research tools."

The next section of this paper discusses the use and potential effects on scientific research of formal intellectual property rights covering academic discoveries. We describe our data and its relevance to this question in Section 3. We then explain our methodology for constructing publication–patent matches, and describe how we construct a plausible counterfactual for our treatment observations in our empirical analysis (Section 4). We present the results of our analysis in Section 5. Finally, we discuss the implications of our results and conclude with a summary of the contributions and limitations of this paper in Section 6.

2 Research and Intellectual Property in Academia

Universities have long been important performers of research, particularly basic research, in the United States and other industrial economies, ranging from 48.9% of U.S. basic R&D expenditures in 1979 to 53.4% in 2009 (National Science Foundation 2012; Table 4-4). At least since the 1970s, this expanded role of U.S. universities in research performance has coincided with growth in patenting and licensing of university discoveries, particularly in biotechnology.

The roles of universities as sources of basic knowledge and as sources of potentially valuable ingredients for commercial innovation raises the possibility of conflict between these roles, with detrimental consequences for the advance of scientific research. Do patents and licenses restrict access to such knowledge? Or is the existence of a patent and/or a license for that patent a signal of the quality of scientific work that leads to greater exploration of the area?

Our examination of the effects of patenting and licensing on scientific communication focuses on "patent–linked publications" (Murray 2002). These are discoveries that are published in scientific journals and become the subject of successful patent applications.[1]

2.1 The Effects of Intellectual Property Protection and Licensing on Scientific Research and Communication

An array of factors, including the Bayh–Dole Act of 1980, other changes in U.S. intellectual property laws and policies, and expanded federal support for academic biomedical research, has increased the patenting of academic research by U.S. universities during and after the 1970s. The growth in such patenting has been the subject of a large literature and considerable debate over its effects on the scientific research enterprise (e.g., Mowery et al. (2004)).

Heller and Eisenberg (1998) argue that expanded patenting of academic research results may result in fragmented and overlapping property rights covering upstream biomedical research, limiting the ability of scientists to access patented and licensed research outputs for follow–on research. Other scholars raising concerns over the expanded assertion of property rights in science include Nelson (2004) and David (2004). Empirical research seeking to assess the effects of patenting on scientific

[1] In our empirical analysis, we allow for the possibility that multiple scientific publications may be associated with each patented discovery, unlike Murray (2002), Murray and Stern (2007), and Fehder et al. (2014), who identify a single patent linked to each publication in their samples. We discuss the construction of our sample of patent–linked publications in Section 4.

communication has examined the effects of patenting on biomedical researchers' willingness to share information on their work (Blumenthal, et al., 1997; Campbell, et al., 2002). More recent research has analyzed the effects of patenting biomedical discoveries that are also disclosed in scientific papers. Some of this work finds that the issuance of a patent results in modest but significant declines in citations to the research papers related to the patent (Murray and Stern 2007, Sampat 2005).[2] Similarly, Williams (2013) finds that patenting of genes by the private firm Celera reduced citations to related scientific research. Other work, however, argues that biomedical researchers rarely if ever search to determine whether a prospective research project or experiment will infringe on patents (Lei et al. 2009, Walsh et al. 2007), raising a question about the mechanism behind any observed citation decreases.

Empirical research on the effects on scientific communication of academic patenting and licensing of discoveries has focused mainly on patenting of academic research results. The effects of university licensing of patents on scientific research have received much less attention from scholars. Unlike patents, licenses are not published or otherwise subject to mandatory disclosure. In many cases the identity of licensees is treated by university technology transfer offices as confidential (Ziedonis 2007).

Why and how might licenses for a specific research advance affect the behavior of academic researchers not involved as authors or patent holders? Sampat and Ziedonis (2005) examined patent citations to Columbia University and University of California patents that were licensed. They find that higher numbers of citations were associated with an increase in the likelihood that a patent would be licensed. Moreover, most citations occurred after the patent was licensed. These scholars interpret this pattern of increased patent citations as indicating market interest in the technological area surrounding the licensed patents. More recently, Drivas et al. (2014) find that citations by non–licensees to patents exclusively licensed (based on restrictions covering either geographic area or field of use) by the University of California increased after the licenses were executed. Similar to Sampat and Ziedonis, Drivas et al. regard the increase in non–licensee patent citations as a reaction to the potential commercial value signaled by the negotiation of the license.

It is possible that a similar signaling effect associated with the execution of a license could in-

[2]In follow–on work, Fehder et al. (2014) find that the negative effect is most pronounced for patent–related papers published in the early years of a journal's founding.

crease citations to patented publications linked to the license. In such a case, the issue of a license "demonstrates" that a particular area of research has scientific value, leading other investigators to pursue work in closely related fields. It is also possible (see Larsen (2011)) that contemporary academic researchers may choose research areas partly based on their potential commercial value, and therefore might respond positively to a "signal" that a given area of research has attracted the attention of industrial licensees. Regardless of whether a license signal operates through perceptions among researchers of scientific or commercial potential, this argument predicts an increase in citations to patented publications following the negotiation of the license.

Equally plausible arguments, however, suggest a chilling effect of licensing on scientific communication. Reactions by university technology licensing offices and/or their licensees to any evidence of patent infringement (even for research purposes, inasmuch as the legal status of the informal research exemption from such infringement suits remains unclear) may be swifter and stronger in the case of patents that are licensed. And licenses may include provisions for reach–through royalties and limitations on the disposition of intellectual property on follow–on research. Moreover, the negotiation of a license may take considerable time, delaying access to the materials or tools embodied in the disclosure.

We are thus agnostic on the likely direction of any effect of licenses on scientific communication associated with publications linked to licensed academic patents. Indeed, both effects may be present for papers in various fields of research, and we hope that our findings shed light on the magnitude of any offsetting effects.

2.2 Research Inputs and Material Transfer Agreements

Patents and licenses increase the "excludability" of intellectual property for other researchers, exposing them to potential legal liability in the event that they utilize or exploit the intellectual property protected by patents and issued to others for their own research or commercial use. A very different form of excludability, highlighted by Walsh et al. (2007), concerns the denial by one researcher of physical access to materials (or other research results) that are inputs to the experiments of another researcher. The survey results of Walsh et al. (2007) indicate that such denials can impose significant costs and delays on the scientific work of other researchers, costs and delays that according to these authors, exceed those associated with patents.

Research inputs have been widely (and imprecisely) identified as "research tools." For example, the NIH Working Group on Research Tools (1998) defines them as "the full range of resources that scientists use in the laboratory... the term may thus include cell lines, monoclonal antibodies, reagents, animal models, growth factors, combinatorial chemistry libraries, drugs and drug targets, clones and cloning tools (such as PCR), methods, laboratory equipment and machines, databases and computer software."[3]

As we noted above, Walsh et al. (2007) argue that denials of access to such research tools are more likely when the erstwhile supplier of them is engaged in "commercial activity," such as licensing of the invention disclosure associated with the tools. Interviews conducted for this study support this view, with one scientist involved in a start–up firm saying "if another company asked to use our [materials] for [the same purpose as our company uses them] we would say 'no.'"[4]

Even when permission is granted, however, researchers gaining access to research tools that are associated with licensed disclosures may encounter difficulties, as highlighted by the celebrated case of the Oncomouse (Murray 2010):

> "In 1984, scientists at Harvard University carefully engineered a new mouse to have a predisposition to cancer, the Oncomouse... The Harvard researchers ... patented their creation and subsequently licensed this patent to DuPont...
>
> [Dupont] set a high price per mouse... placed restrictions on breeding programs... demanded publication oversight... [and] insisted upon a share of any commercial breakthroughs made using the Oncomouse."

These and similar restrictions may have an adverse effect on the use by other researchers of research tools such as the Oncomouse. Moreover, as the NIH Working Group on Research Tools (1998) noted in its report, licensees may have an incentive to restrict access to these materials:

> "If the sponsor or licensee plans to develop the research tool as a commercial product for sale to researchers, it may be unwilling to permit the university to undercut its position in this particular market by giving the tool away to potential paying customers."

Surveys of scientists by Walsh et al. (2007) and Lei et al. (2009) find that requests by researchers for research tools from industrial researchers, a group more likely to be engaged in commercial applications of research, were rejected approximately twice as often as requests to other academics.

[3]The National Research Council panel has a similarly broad definition that includes materials that "... may be critical inputs for the success of a research project." (National Research Council 2010; p. 7).

[4]Even though this scientist was unwilling to share materials in this instance, in other instances he/she had shared materials.

As the discussion of "research tools" indicates, developing a definition that facilitates their identification and empirical analysis is challenging. The definitions of research tools employed by the NIH Working Group and Walsh et al. (2007) are very broad and do not lend themselves to empirical operationalization. Instead of attempting to develop and defend a definition of research tools that relies on the characteristics of the relevant invention disclosure or patent, we identify research tools based on the existence of a Material Transfer Agreement (MTA) associated with a patented invention disclosure.

Material Transfer Agreements (MTAs) are agreements that govern the transfer and exchange of materials, usually biological, used in research. Although the informal exchange by researchers of biological materials for use in fundamental research has a long and occasionally controversial history in the biomedical sciences, these materials exchanges historically were governed by little more than a letter from the source accompanying the materials, requesting acknowledgement and in some cases asking that the materials not be passed on to third parties (McCain 1991). The more elaborate MTAs used in contemporary materials exchanges appear to be a byproduct of the post–1980 surge in academic patenting (Streitz and Bennett 2003).

One of the few analyses of the role of MTAs in the scientific research enterprise is Scott Stern's discussion of biological resource centers (Stern 2004). Biological resource centers (BRCs) are nonprofit materials depositories that play a key role in maintaining the reliability and provenance of cell lines used by industrial and academic researchers—as Stern notes, contamination of widely used cell lines has caused major research fiascoes in the past several decades. Stern argues that the use of MTAs by BRCs has aided the exchange of materials, and recommends that MTAs be a standard complement to patents covering biological discoveries: "Putting MTAs in place at the time of patent approval lowers the cost of mutually beneficial transactions between the developers of materials and follow–on researchers and widens the availability of patented biomaterials" (2004; pp. 96–97). Similarly, Walsh et al. (2003) argue that the formalization of materials exchanges through MTAs may simplify these transactions and facilitate researcher access.

To confirm that MTAs are a good indicator that an invention disclosure is associated with a research tool, we examine UC Berkeley data on incoming MTAs—that is, the agreements accompanying research materials requested from other researchers by UC Berkeley researchers. These data describe the requested materials and their intended use by the recipient UC Berkeley re-

searcher. We analyzed a random sample of 50 of these MTAs, and found that 44% were related to DNA/RNA/Plasmids, 32% concerned cell lines or other biological/chemical agents, 16% were animal models, 6% were data transfers, and 2% were concerned with "other non–research inputs." Overall, therefore, 98% of these MTAs involved materials that fit within the NIH Research Tools Working Group definition of research tools. We also analyzed the intended use for the materials requested through the MTA by UC Berkeley researchers, and found that 94% of the MTAs indicated that the requested material was to be used as an input to further research and a further 4% implicitly indicated that such uses were intended.

This analysis of UC Berkeley MTAs leads us to conclude that our treatment of the presence of an MTA as an empirical indicator that a given invention disclosure is a research tool and/or input to follow–on scientific research is defensible. It is important nevertheless to note two caveats associated with our empirical use of MTAs as indicators of research tools. As Walsh et al. (2007) and Mowery and Ziedonis (2006) point out in their discussions of MTAs, a majority of the materials transfers among academic scientists do not rely on formal MTAs that are disclosed to academic Technology Transfer Offices (TTOs). Although we believe that the presence of an MTA is a reasonable indicator that a given disclosure has applications as a research tool, our data in fact contain many other disclosures (including disclosures that are patented and licensed) that may well be research tools but (lacking an MTA) cannot be identified as such. It is likely that our empirical approach thus understates the effects on citations of licensing of patented disclosures that are research tools. In addition, our data enable us to only identify the "effects of MTAs" that are negotiated and agreed to by all parties to the materials transfer. In other words, and in contrast to Walsh et al. (2007), we do not identify the effects on scientific research of the denial by researchers of other researchers' requests for research materials.

Based on these arguments, we anticipate that access to research tools (inputs to other scientific experiments) that are associated with licensed intellectual property (a sign of the commercial exploitation, prospective or otherwise, of the intellectual property) may well be restricted, even when terms for its exchange among researchers are successfully negotiated through a Material Transfer Agreement (MTA). Licensing of IP related to research tools thus may have negative consequences for follow–on scientific research and therefore may have a negative effect on citation rates for publications related to such IP.

3 Data

We draw on two principal sources of data for our empirical analysis. The first, the "IP data," is an extract from the technology disclosure database maintained by the technology transfer office within the University of California Office of the President (UCOP). UCOP monitors and in some cases manages invention disclosures, patent applications, and licensing transactions for all campuses of the University of California (nine campuses, including five medical schools, during the period of our study).

These data list all 11,341 inventions reported by University of California faculty from 1997 to 2007. These disclosures led to 2,035 issued U.S. patents, 1,890 licenses to these patents, and 3,853 MTAs by the end of 2009. Note that only a small subset of technology disclosures is patented, and universities' patenting propensity varies among fields of academic research—since the 1980s, patenting and licensing activity at UC has been dominated by biomedical research (Mowery et al. 2004). The distribution of MTAs also is highly skewed, with few disclosures generating the majority of MTAs and many disclosures associated with no MTAs. This echoes the finding in Mowery and Ziedonis (2006) that MTAs are disproportionately concentrated in biomedical fields of research, as are licenses.

The second source, "publications data," comes from Web of Science, an internet–based service that tracks the bibliographic information and the citations to and from articles published in 10,000 of the highest-impact journals across 256 disciplines.[5] This database supplied the title, author names, journal, publication date, and citation information for each scientific paper. The information on "forward citations," citations from later published papers to that publication, was extracted through the end of 2009. The Web of Science also provides a number of well–accepted measures of journal quality. The most prominent of these is an "impact factor" that measures the average number of times an article in that journal is cited in its first two years, and which we include in our analysis.

[5] `http://thomsonreuters.com/products\_services/science/science\_products/a--z/web\_of\_science` (downloaded May 2010).

4 Methodology

In this section we describe our empirical methodology. We explain how we identify the publications related to each patent, how we construct groups of treatment and control observations that are comparable save for the treatment effect, and finally we discuss our estimation technique.

Our analysis searches for differences in the citations associated with publications linked to patented discoveries, depending on whether or not they are licensed. One advantage of restricting our comparison set to patent–linked publications is that they are likely to be similar in quality and other characteristics. In particular, since all of the underlying disclosures are patented, differences in "commercializability" are considerably lower than they would be in a general sample. Nevertheless, there may well exist other unobserved differences between the patented publications that are licensed and those that are not. Below we describe how we construct a control sample and employ an difference–in–differences specification to address this issue.

4.1 Linking Invention Disclosures, Patents, and Scientific Publications

Observing the effect of licensing a patent on citations to a related publication requires identifying the linkages among publications, patents, and licenses. The connections among disclosures, patents, and licenses are contained in the data provided by UCOP, which tracks these directly. But the UC data themselves provide no direct link between published papers and any patents covering the research advances covered by them–what Murray and Stern (2007) terms "patent-paper pairs."

Prior efforts to link patents to scientific publications used techniques that differ somewhat from those developed in this paper. Murray and Stern (2007) matches patents to articles published in *Nature Biotechnology* by reading both patents and the academic articles and relying on expert judgment to connect them. Azoulay et al. (2012) used an algorithm that matches specific characteristics of publications in the biomedical sciences that are included in the PubMed database to information contained in the scientific references of patents within biomedical and chemical patent classes to identify patent–paper pairs for 9,483 academic scientists working in the life sciences.

The Murray and Stern (2007) and Azoulay et al. (2012) techniques are less well suited to our study. The size and diversity of our sample makes it difficult for us to assemble the requisite scientific and technical expertise to assess each potential patent–paper pair by hand. In addition,

our construction of this dataset matching patents and scientific papers covers a range of scientific disciplines that includes not only the life sciences but also physical sciences, which means that the PubMed database would not cover our entire set of publications.[6]

Our methodology instead employs an "inventor–based matching" technique to link patents to scientific papers. Inventor–based matching relies on two assumptions. First, inventors listed on a patent are likely to be the authors listed on related publications. Second, the patent application date is likely to occur near the publication date of the academic article. Based on these assumptions we construct a maximum–likelihood estimator for the publication(s) that best matches a particular patent. For each inventor name listed in the patent, we first identify all publications authored within a five–year window that includes the year of the patent application and the two years prior to and following the application year (*t-2* to *t+2*). We match papers to patents by selecting the publications for which the inventors listed on the patent overlap to the greatest extent with the co–authors of the papers that are linked to the patent. Those publications with the greatest overlap are chosen as matches. For example, consider U.S. patent number 7,011,723, "Adhesive Microstructure and Method of Forming Same." This patent pertains to adhesives inspired by the physiology of the foot of the gecko and credits four inventors. In this case, we extract four publication sets (one for each inventor) and retrieve publications common to all four inventors. Figure 1 illustrates this process for this patent.

*** Figure 1 Here ***

Our approach can result in multiple publications as "best" matches, in contrast to Murray and Stern (2007), which links one publication to each patent. For the example patent described above, the inventor–based matching approach yields two matches of the underlying invention to publications in scientific journals. The first, entitled "Adhesive Force of a Single Gecko Foot–Hair," appeared in *Nature* in March 2000. A second article, "Evidence for Van der Waals Adhesion in Gecko Setae" was published in *Proceedings of the National Academy of Sciences* in 2002. In this case, all four patent inventors were listed on each publication.

Inventor–based matching does not restrict us to instances in which all inventors are listed as authors on the publication. For instance, if in the example above a lab technician was also listed

[6]Non-biomedical fields are well represented in our data as roughly half (52%) of our observations come from outside the disciplines of biochemistry, biology, and medicine.

as an inventor on the patent, but was not included on any linked academic publications, the algorithm would choose the publication(s) with the maximum possible overlap. In this instance the publications listing four of the five inventors would be chosen since there would be no five–out–of–five–inventor matches.[7]

As highlighted above, the inventor–based matching method enables us to link a single patent with more than one publication. Such a match will occur precisely when multiple publications share the same level of overlap between the inventors, and no publications have a greater overlap.[8] In general, accurate matches are less likely when the precision of the estimate is low. For example, our matching algorithm would be unlikely to produce correct matches if, on a four–inventor patent, we identified only publications that listed a single inventor as an author. In this case, the algorithm would identify as matches all publications by all of the inventors in the relevant (*t-2*, *t+2*) window, many of which could be false positives (papers where an author has the same name as the patent inventor but is not the same individual). To improve the precision of the estimate and limit such Type I errors, we restrict matches to those with three or more inventors/authors. The logic behind this criterion is illustrated in Figure 2, which plots the number of papers matched to each patent in our dataset linked by at least two inventors and authors and lists the number of names common to both the patent and the published paper.

*** Figure 2 Here ***

Figure 2 shows that 82% of the sample patents are linked by the inventor–based matching algorithm to between one and five papers, while the remaining 18% of our patents match to six or more publications. The large number of papers associated with each of the patents in the 18% may reflect common scientist names (e.g., "J. Smith"). Figure 2 also demonstrates that patents matched

[7] Based on a survey of life scientists, Haeussler and Sauermann (2013) propose that the underlying processes determining publication authorship and patent inventorship differ. There is little evidence to suggest that these differing processes would bias our inventor–based matching methodology, however.

[8] We assume that a publication and a patent are more likely to be a match if they share an author:

$$p\left(match_{pub_i,patent_j} | \, author_k \in \left(authors_{pub_i} \cap inventors_{patent_j}\right)\right) > p\left(match_{pub_i,patent_j}\right)$$

Here $authors_{pub_i}$ and $inventors_{patent_j}$ are the sets of authors for publication i and the inventors for patent j, respectively. Given these parameters, pub_m is a "match" for $patent_j$ if

$$m \in \underset{i}{\text{argmax}} \prod_{i,k} p\left(match_{pub_i,patent_j} \mid author_k \in \left(authors_{pub_i} \cap inventors_{patent_j}\right)\right)$$

As with all maximum–likelihood estimators, a "best" estimate is not necessarily precise (Casella and Berger 2002).

to more than five publications typically have only two names that are common to both the inventor list and the author list.

The restriction of three or more inventors/authors imposes the following conditions on our sample: (a) a patent must have at least three inventors, and (b) the associated publication must list at least three of those inventors as authors. Figure 3 depicts the consequence of these restrictions on the sample.

*** Figure 3 Here ***

Column 1 in Figure 3 represents the 2,035 patents in the IP data. By excluding patents with fewer than three inventors, we omit 944 patents listing one or two inventors (Column 2) from the full sample, leaving 1,091 patents. From these 1,091 we exclude an additional 363 three–or–more–inventor patents where fewer than three inventors were listed as authors on any publication, resulting in a remaining sample of 728 patents (Column 3). Of these 728 patents, 406 list three inventors, 201 list four inventors, and 121 list 5 or more inventors. The fourth column of Figure 3 reports the number of journal citations "per patent," (i.e., the number of journal citations for all publications that are matched, using the three–name overlap restriction, to that patent).

We examine the effect of restricting our sample to higher levels of inventor–author overlap by comparing the sample statistics of 3–inventor overlap and 4–inventor overlap samples, as shown in Columns 1 and 2 in Table 1.

*** Table 1 Here ***

The average publication year, publication age when the citations are observed, and the publication ages when licenses and MTAs are issued are stable across the samples, thus offering little reason to expect that differences in means for these variables will affect the results of our statistical analysis. In contrast, two measures of publication quality, the number of citations per year and the average impact factor of the publication's journal, are higher in the 4+ inventor overlap sample in Column 2. The 4+ inventor overlap restriction thus appears to produce an increase in average publication quality, which we argue supports our usage of higher overlaps to reduce "false positives"—papers with a set of co–authors that matches those on a patent, but which are not actually linked to the patent. Since Murray and Stern (2007) report that publications linked to patents

receive more citations than publications with no associated patents, the exclusion of false matches (i.e., unpatented publications) should increase publication quality—which is what we observe.

To assess the validity of this matching algorithm, we compare its output with the hand–matches compiled by Murray and Stern (2007).[9] For each patent in the Murray and Stern sample, we generated maximum–likelihood estimates of the "best" publication matches using our inventor–based matching technique.[10] Of the 170 patent–publication pairs identified by Murray and Stern, our automated method determined an identical "best" publication match for 95% of the sample. In a small number of cases (4%), our algorithm identified a "better" match.[11] Only in two cases (1% of their sample) did our approach yield matches otherwise inconsistent with their hand–matching process.

In summary, our inventor–based matching approach, while demonstrating accuracy comparable to the hand–matching approach employed by Murray and Stern (2007), possesses several advantages: (a) it does not impose a simple one–to–one relationship between patents and publications; (b) it is transparent, reproducible, and does not rely on domain expertise; (c) it is automated; and (d) it is generalizable across scientific fields.

4.2 Construction of Treatment and Control Groups

Designing an adequate specification to test the effect of licensing on scientific communication is a challenging exercise, because of the complex and varied patterns of citations that publications may receive during our sample period. Citations to some publications may grow throughout the period covered by our data, whereas for other publications, citations may rise, peak, and then decline. Moreover, the timing and rate of any ascent and descent in citations may vary. It is thus difficult to construct a sufficiently flexible parametric model to adequately accommodate these differences.[12]

[9] Fiona Murray and Scott Stern graciously provided their data to us for comparison.

[10] We used a 2+ inventor overlap sample in this test to ensure that we use as much of this limited hand–matched sample as possible for validating our method. The 3+ overlap required for our main analysis provides stronger evidence for matches than what we test here.

[11] The difference here is likely because of the direction of matching. Murray and Stern began with a set of publications and found the most–similar patent, whereas our analysis starts with a patent and finds the most–similar publications. The patent identified by Murray and Stern may have been the best match for that publication, but another publication may be an even better match to that patent.

[12] For example, it is reasonable to assume that publications in more highly cited journals accrue more citations, which would argue for including *Journal Impact Factor* as a control variable. Similarly, the academic discipline (hereafter *Journal Subject*) could also drive citation patterns, as could patterns of citations to a publication prior to the license (*Citations in t-1*, *Citations in t-2*). Each of these the effects could be non–linear, which would suggest the inclusion of higher–order terms. Interaction terms among these variables would also be important, since the

Because of the complex and potentially non–linear interactions of various control variables that we employ to construct treatment and matching control groups, we pursue a flexible nonparametric approach. This technique allows us to weaken the assumption of linearity and enables us to account for the many plausible interaction effects that could be present in our analysis.

The non–parametric method we use is "nearest neighbor" matching. It searches the set of non–treatment observations to identify the one "closest" to each treated observation. Collectively, these "closest" non–treatment observations form a control group. Because this search is done based on observable characteristics, the control and treatment observations should be similar along these dimensions.[13]

We identify these "nearest neighbors" using the "Matching" package for `R` software,[14] which employs the evolutionary search algorithm "genetic matching" to pair treatment observations with potential controls and iteratively improve covariate balance (Diamond and Sekhon 2013, Sekhon 2011).[15] According to Sekhon (2011), genetic matching "dominates the other matching methods in terms of MSE [Mean Squared Error] when assumptions required for EPBR [Equal Percent Bias Reduction] hold and, even more so, when they do not."[16]

Our procedure employs two types of variables, those where we specify an exact match and those where a nearby (nearest neighbor) match is sufficient. We require an exact match on the following variables:

- Publication Age: Number of years between paper publication date and year of "treatment," i.e., license issue
- Journal Subject: Academic discipline of the journal (e.g., medicine)

effects of *Journal Impact Factor* could differ by discipline, to cite only one example. Each of these effects could affect a publication differently at different points in time, calling for them to be interacted with age fixed effects. Fully interacting all of these effects would result in many of the parameter estimates being determined by small numbers of observations, or lacking observations at all.

[13]Restricting ourselves to control observations that satisfy our matching criteria decreases our sample size. The effects of a smaller sample size on the precision of our estimates, however, are ambiguous. Smaller samples will tend to reduce our precision (making the standard errors larger), but this loss of precision may be offset by greater homogeneity within the smaller sample of treatment and control observations.

[14]Available from the Comprehensive `R` Archive network at `http://CRAN.R-project.org/package=matching` (R Core Team 2014)

[15]Treatment and control variables are in "covariate balance" when they have the same joint distribution of their observed covariates (Diamond and Sekhon 2013).

[16]This is the first innovation study of which we are aware that employs the nearest neighbor technique with genetic matching in place of other procedures such as propensity scoring or coarsened exact matching. Previous studies in the social sciences that utilize genetic matching include Morgan and Harding (2006), Gilligan and Sergenti (2008), Eggers and Mainmueller (2009), Ladd and Lenz (2009), Gordon (2009), and Hopkins (2010).

- Patent Granted (yes/no): Whether the related patent has been granted at the time of license
- MTA Issued (yes/no): Whether the paper has an associated MTA at the time of the license

For example, these restrictions imply that in our analysis a (licensed) treatment observation in the life sciences with an issued patent and no MTA would be compared with a non–licensed control group observation in the life sciences with an issued patent and no MTA, and the number of years since publication would be identical for the treatment and control observations.

For each treatment observation that matches on these exact characteristics, we then choose its nearest neighbor based on its proximity in the following five characteristics:

- Journal Impact Factor
- Publication Year
- Citations in *t-2*: Number of citations two years before the treatment
- Citations in *t-1*: Number of citations one year before the treatment
- Slope of citation curve between *t-2* and *t-1*.

In each of these dimensions we limit the maximum "distance" between each treatment and control observation to one standard deviation for that variable (the "caliper"), thus excluding any observation that differs by more than that amount for any characteristic.[17] Under this procedure, treatment observations with no equivalent control observation are dropped from the sample. Collectively, these restrictions produce control observations such that for each characteristic, the control matches either exactly or within one standard deviation to the corresponding treatment observation.

Before we estimate the effects of licensing on citations to patented publications, we must verify the covariate balance between the treatment and control groups. An effective matching procedure should yield summary statistics for the treatment and control groups that are similar. Table 2 reports the mean of each variable for the treatment and control groups produced by genetic matching and the results of t–tests (difference of means) and Kolmogorov–Smirnoff (KS) tests (difference in distributions) between these two groups.

*** Table 2 Here ***

[17] Coarsened exact matching is one alternative matching approach. Employing a caliper–based method rather than coarsened exact matching, however, allows us to exclude observations whose observable covariates would make them outliers as well as those which would qualify them as "inliers," that is, observations that are within the range of the data, but nevertheless lack a comparable control observation (Sekhon 2011).

The means and distributions for first four variables reported in Table 2, *Publication Age*, *Journal Subject*, *Patent Issued*, and *MTA Issued*, are equal between the treatment and control groups, which is not surprising in light of our requirement for an exact match in these dimensions across the two groups. For the remaining variables, means for the control and treatment groups are similar, although there are economically small, but statistically significant differences for *Journal Impact Factor*, *Citation Slope* and *Citations in t-1*. These statistically significant differences suggest that further analysis (in our case: a difference–in-differences specification) should be used to ensure that these differences are not driving the result. That said, the small magnitude of these differences suggest that these variables would have to have very large coefficients to meaningfully bias our results (e.g., a *Journal Impact Factor* difference would have to cause a very large difference in post–licensing citations). The analysis described below confirms that these differences do not meaningfully influence our estimates.

4.3 Estimation

Once we have generated our treatment and control group sample, we employ a difference–in–differences approach to account for any remaining unobserved differences between the treated and control observations and to estimate the size and direction of the treatment effect. More specifically, we compare the change in the number of citations to one patented publication following the execution of a license (a treated observation) to the change in number of citations for a comparable publication that lacks a license (a matching control observation). A citation by a scientist to his or own prior publication would not represent a knowledge flow, thus we exclude papers where at least one of the authors of the citing publication is also an author of the cited publication (i.e., a self–citation).[18] We define our outcome of interest as:

$$\begin{aligned} Treatment\ Effect_{(t-1)\rightarrow(t+i)} = &(Citations_{t+i} - Citations_{t-1})_{pub_{w/\ License}} \\ &- (Citations_{t+i} - Citations_{t-1})_{pub_{w/o\ License}} \end{aligned} \tag{1}$$

where i is the number of years after the license, ranging from 1 to 3.

[18] A few publications receive large numbers of citations, creating a concern that these outliers could unduly influence our empirical results. To ensure that the results reflect the central tendency of the data rather than these extremes, we trim the 2.5% highest and lowest values and estimate our coefficients on the remaining 95% of the sample.

The specification is as follows:

$$\begin{aligned}\Delta Citations_{(t-1)\rightarrow(t+i)} = \beta_0 &+ \gamma License + \beta_1 Age \\ &+ \beta_2 Publication\,Year + \beta_3 Journal\ Subject + \beta_4 Journal\ Impact\ Factor \\ &+ \beta_5 Citations_{t-2} + \beta_6 Citations_{t-1} + \beta_7 Citations\ Slope \\ &+ \beta_8 MTA\ Issued + \beta_9 Patent\ Granted + \epsilon\end{aligned} \tag{2}$$

Our coefficient of interest is γ, which we present in our results.[19]

Using a difference–in–differences estimator allows us to avoid bias associated with changes that affect the "before" and "after" periods equally. For example, in the matched sample case, if a particular publication receives on average five additional citations per year due to some unobserved covariate, these citations will be included in both $Citations_{t-1}$ and $Citations_{t+1}$ terms, thus the impact on the estimate will be zero.

To summarize, our analysis employs two techniques. First, we identify the nearest neighbor to each of our treatment observations using genetic matching. This non–parametric matching technique means that we make fewer assumptions about the parametric form of the effect than we would with a linear or generalized linear (e.g., negative binomial). Based on this sample construction, our results should be interpreted as an Average Treatment Effect on the Treated (ATT). In other words, our sample construction implies that our estimates of the effects of licensing on citations have the most external validity for patent–linked publications whose covariates resemble those that receive licenses.

[19] Because we expect the differences in covariates between the treated and control groups after the nearest neighbor matching to be smaller, the linearity assumption embedded in least–squares is more plausible than it would be for an unmatched sample. This assumption nonetheless introduces a lack of flexibility into the specification. Rubin (1979) discusses the value of using these (slightly modified) techniques and concludes that "pair–matching coupled with regression adjustment on the matched pairs is a quite effective general plan for controlling the bias due to matching variables, and this combination is clearly superior to regression adjustment" (p. 318).

5 Results

5.1 The Effects of Licensing on Citations to Patent-Linked Publications

We now turn to estimating the effects of licensing on scientific communication as measured by citations to patent-linked publications. We report the results of our findings in two ways. First, the effect can be seen directly through the citation pattern for the treatment group (dashed line with circles) and the control group (dotted line with squares), as shown in Figure 4.

*** Figure 4 Here ***

As depicted in Figure 4, patent-linked publications that are licensed receive more citations commencing two years after a license is executed than do publications that are linked to unlicensed patents. This pattern of increased citations two years after the license is consistent with a more gradual expansion of awareness within the research community of the license that increases citations to the relevant publication only after a lag, reflecting the lack of any public announcement of the license. In this interpretation, scientists gradually adjust their research agendas to intensify work in the area of research covered by the license, in response to the positive signal associated with the license of the commercial or scientific quality of related research.

Figure 5 reports results from the difference–in–differences analysis (Equation 2), confirming the post–licensing citation pattern presented in Figure 4 and testing for the statistical significance of the differences in citations between the two groups. The effect of licensing on the number of citations is near zero and not statistically significant in the first year after the execution of the license (γ = -0.20 with a standard error of 0.51). The average difference in citations between the treatment and control groups is 1.72 additional citations received by a paper linked to a licensed patent in the second year (standard error of 0.68) and 1.68 citations received by such a paper in year *t+3* (standard error of 0.77), both of which are significant at the 1% level. This pattern is very similar to the citation trends depicted in Figure 4 due to the good covariate balance prior to the diff–in–diffs regression. The magnitude of the differences between the two groups in years *t+2* and *t+3* implies that the average publication receives an increase of approximately 25% in citations in these two years compared to the control group.

*** Figure 5 Here ***

Inspection of the residuals after our regressions (not reported) shows that the mean and median are aligned and that the residuals are distributed approximately normally, supporting the validity of our model.

5.2 The Effects of Licensing on Citations to Patented Publications Associated with Research Tools

We now turn our attention to the effects of licensing on patented publications that we believe are more likely to be associated with research tools. In their test of the effects of patenting on citations to "research tool"–related publications, Murray and Stern (2007) defined such publications as those linked to any patents within their sample in the 3–digit patent classes 435 (Chemistry: Molecular Biology and Microbiology) and 800 (Multicellular Living Organisms and Unmodified Parts Thereof and Related Processes). A patent class–based definition, however, does not account for the patent's use. Instead of employing patent classes, as we noted above, we believe that patented publications for which MTAs are issued are likely to fall within a rough definition of "research tool."

Our analysis of the effects of licensing on this class of patented publications restricts the sample to only those with material transfer agreements.[20] Thus, both treatment and the control observations have MTAs and the difference between them is whether a license is issued.[21] Table 3 presents the covariate balance for this sample.

*** Table 3 Here ***

Although the difference in means for the control and treatment groups in the MTA–linked sample is larger in absolute magnitude than that for the full sample reported in Table 2, the MTA sample reveals fewer statistically significant differences in means. For the MTA–linked sample, only the t–test of the difference of means for *Publication Year* is highly significant (at the 1% level),

[20] We include MTAs that are issued before or after the license. In doing so we assume that an invention receiving an MTA irrespective of when the MTA was executed (i.e., the execution of the MTA) does not "convert" the material into a research tool. Restricting the sample to instances where an MTA exists prior to the license could ensure against possible reverse causality, however. This restriction led to results that are similar in direction, although with much smaller sample sizes. Thus we report results with the larger sample of MTA–linked inventions.

[21] Because we also match treatment to control observations based on whether the observation has an MTA, treated observations with MTAs prior to license are matched to controls that also already have an MTA, while those that have not received an MTA prior to license are matched to controls that also have not yet received an MTA as of the license execution date.

while the difference in distributions between the treatment and control group observations is weakly significant (10%) only for *Journal Impact Factor*.

As we did in our analysis for the full sample, we present both the non–parametric citation curve for the matched samples in Figure 6 and the regression adjusted results in Figure 7 below.

∗∗∗ Figures 6 and 7 Here ∗∗∗

Our difference–in–differences analysis of the effects of licensing on citations to patent–paper pairs that are associated with MTAs reveals a very different pattern from that observed in the overall sample. For the overall sample, licensing is associated with an increase in citations to the publications linked to the underlying patented disclosure. For the MTA–linked sample, however, licensing is associated with a decline in citations. Using the covariate adjusted values (Figure 7), the magnitude of the licensing effect is -3.37 citations in year *t+1* (standard error of 0.79) and -4.38 in year *t+3* (standard error of 1.17). The differences in these two years are statistically significant at the 1% level, but is not statistically significant from zero in year *t+2* (γ = -0.71 with a standard error of 0.81).[22] Overall, these coefficients represent a decrease in citations of approximately 40%–50% for the average publication, although our small sample size calls for caution in interpreting the exact magnitude of the effect.

The timing of the "license effect" for citations to MTA–linked patent–paper pairs also suggests a more rapid negative impact than was observed for the positive effect of licenses on citations in the overall sample. The apparently more rapid appearance of the negative effects on citations observed in the MTA sample is broadly consistent with the delays and project abandonment observed in the surveys by Walsh et al. (2007) and Lei et al. (2009), as well as interviews we conducted for this study. Moreover, declines in citations for the MTA–linked disclosures that are licensed occur as early as the year of the execution of the license. Inasmuch as the citations to the focal publication associated with the patented disclosure appear only after a lag, the speed of this observed effect suggests that researchers or universities may limit or impose other restrictions on sharing of "research tools" during the negotiation of the license.[23]

[22] We conducted a robustness check using a 2–inventor overlap sample which resulted in coefficients of -4.9***, -4.7***, and -7.1*** for years *t+1*, *t+2*, and *t+3*). Based on these results we do not ascribe any economic meaning to the unusual result for *t+2* in the original sample.

[23] Note that the identification of any in–year effect is weaker, since the exact timing of the license during the year is not accounted for in the analysis.

6 Conclusion

This paper has investigated the effects of licenses on communication among scientists of published research results that are linked to patented academic invention disclosures within the University of California system. Our results suggest that in general, licenses on scientific work are associated with an increase in the number of citations to related publications, but that this effect differs for a class of patented invention disclosures that we believe includes a high share of research tools. For these inventions we observe a decrease in the number of citations following a license, suggesting that the benefits of MTAs for access to research tools noted by Stern (2004) or Walsh et al. (2003) in their discussions of MTAs may be limited or absent in this sample.

Our results are consistent with other findings in the literature that suggest that licensing may have a positive signaling effect (e.g., Drivas et al. (2014)), but that licenses on research tools may lead to restrictions on input materials that are important for follow–on research (Heller and Eisenberg 1998, Walsh et al. 2007).

These findings should be interpreted with caution, however. Our sample of patented publications represents discoveries made by scientists at a single institution. While the University of California is the most prolifically patenting university in the U.S. according to a recent study by the U.S. Patent and Trademark Office (2014), this university may not be representative of all academic institutions.

This study makes four contributions. First, it introduces a novel method for linking patents and publications associated with an academic research advance, and thereby automates a previously arduous process that has been difficult to scale up. Second, it operationalizes a definition of research tools that permits empirical analysis of such discoveries. Third, it employs a matching methodology that avoids strong parametric assumptions that may not be appropriate for the complex interactions that underlie journal citation patterns. And finally, but most importantly, this paper contributes to our understanding of the effect of the licensing of intellectual property rights covering academic research output on scientific communication.

References

Azoulay, P., Sampat, B. N., and Zivin, J. S. G. (2012). The Diffusion of Scientific Knowledge Across Time and Space: Evidence from Professional Transitions for the Superstars of Medicine . In Lerner, J. and Stern, S., editors, *The Rate and Direction of Innovative Actvity Revisited*, pages 107–160. Chicago.

Casella, G. and Berger, R. L. (2002). *Statistical Inference*. Pacific Grove, CA, 2nd edition.

David, P. A. (2004). Can "Open Science" be Protected from the Evolving Regime of IPR Protections? *Journal of Institutional and Theoretical Economics*, 160(1):9–34.

Diamond, A. and Sekhon, J. S. (2013). Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies. *Review of Economics and Statistics*, 95(3):932–945.

Drivas, K., Lei, Z., and Wright, B. D. (2014). Academic Patent Licenses: Roadblocks or Signposts for Unlicensed Innovators? *Working Paper*, pages 1–70.

Eggers, A. C. and Mainmueller, J. (2009). MPs for Sale? Returns to Office in Postwar British Politics. *American Political Science Review*, 103(04):513–533.

Fehder, D. C., Murray, F., and Stern, S. (2014). Intellectual Property Rights and the Evolution of Scientific Journals as Knowledge Platforms. *International Journal of Industrial Organization*, Forthcoming:1–45.

Gilligan, M. J. and Sergenti, E. J. (2008). Do UN Interventions Cause Peace? Using Matching to Improve Causal Inference. *Quarterly Journal of Political Science*, 3(2):89–122.

Gordon, S. C. (2009). Assessing Partisan Bias in Federal Public Corruption Prosecutions. *American Political Science Review*, 103(4):534–554.

Haeussler, C. and Sauermann, H. (2013). Credit Where Credit is due? The Impact of Project Contributions and Social Factors on Authorship and Inventorship. *Research Policy*, 42(3):688–703.

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

Hopkins, D. J. (2010). Politicized Places: Explaining Where and When Immigrants Provoke Local Opposition. *American Political Science Review*, 104(1):40–60.

Ladd, J. M. and Lenz, G. S. (2009). Exploiting a Rare Communication Shift to Document the Persuasive Power of the News Media. *American Journal of Political Science*, 53(2):394–410.

Larsen, M. T. (2011). The Implications of Academic Enterprise for Public Science: An Overview of the Empirical Evidence. *Research Policy*, 40(1):6–19.

Lei, Z., Juneja, R., and Wright, B. D. (2009). Patents versus patenting: implications of intellectual property protection for biological research. *Nature Biotechnology*, 27(1):36–40.

McCain, K. W. (1991). Communication, Competition, and Secrecy: The Production and Dissemination of Research-Related Information in Genetics. *Science, Technology, and Human Values*, 16(4):491–516.

Morgan, S. L. and Harding, D. J. (2006). Matching Estimators of Causal Effects: Prospects and Pitfalls in Theory and Practice. *Sociological Methods & Research*, 35(1):3–60.

Mowery, D. C., Nelson, R. R., Sampat, B. N., and Ziedonis, A. A. (2004). *Ivory Tower and Industrial Innovation: University-Industry Technology Transfer Before and After the Bayh-Dole Act in the United States*. Stanford University Press, Stanford, CA.

Mowery, D. C. and Ziedonis, A. A. (2006). Academic Patents and Materials Transfer Agreements: Substitutes or Complements? *The Journal of Technology Transfer*, 32(3):157–172.

Murray, F. (2002). Innovation as Co-Evolution of Scientific and Technological Networksw: Exploring Tissue Engineering. *Research Policy*, 31:1389–1403.

Murray, F. (2010). The Oncomouse that Roared: Hybrid Exchange Strategies as a Source of Distinction at the Boundary of Overlapping Institutions. *The American Journal of Sociology*, 116(2):341–388.

Murray, F. and Stern, S. (2007). Do Formal Intellectual Property Rights Hinder the Free Flow of Scientific Knowledge? *Journal of Economic Behavior and Organization*, 63(4):648–687.

National Institutes of Health Working Group on Research Tools (1998). Report of the National Institutes of Health (NIH) Working Group on Research Tools: Presented to the Advisory Committee to the Director, June 4, 1998. pages 1–56.

National Research Council (2010). *Managing University Intellectual Property in the Public Interest*. The National Academy Press, Washington, D.C.

National Science Foundation (2012). Science and Engineering Indicators.

Nelson, R. R. (2004). The Market Economy, and the Scientific Commons. *Research Policy*, 33(3):455–471.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rubin, D. B. (1979). Using Multivariate Matched Sampling and Regression Adjustment to Control Bias in Observational Studies. *Journal of the American Statistical Association*, 74(366):318–328.

Sampat, B. N. (2005). Genomic Patenting by Academic Researchers: Bad for Science? *Working Paper*, pages 1–48.

Sampat, B. N. and Ziedonis, A. A. (2005). Patent Citations and the Economic Value of Patents. In Moed, H. F., Glänzel, W., and Schmoch, U., editors, *Handbook of Quantitative Science and Technology Research*, pages 277–298. Kluwer Academic Publishers, Dordrecht.

Sekhon, J. S. (2011). Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The Matching package for R. *Journal of Statistical Software*, 42(7):1–52.

Stern, S. (2004). *Biological Resource Centers: Knowledge Hubs for the Life Sciences*. Brookings Institution Press, Washington, D.C.

Streitz, W. D. and Bennett, A. B. (2003). Material Transfer Agreements: A University Perspective. *Plant Physiology*, 133:10–13.

U.S. Patent and Trademark Office (2014). U. S. Colleges and Universities—Utility Patent Grants, Calendar Years 1969–2012. http://www.uspto.gov/web/offices/ac/ido/oeip/taf/univ/org_gr/t250_univ_ag.htm.

Walsh, J. P., Arora, A., and Cohen, W. M. (2003). Effects of Research Tool Patents and Licensing on Biomedical Innovation. In Cohen, W. M. and Merrill, S. A., editors, *Patents in the Knowledge–Based Economy*, pages 285–340. The National Academies Press, Washington, DC.

Walsh, J. P., Cohen, W. M., and Cho, C. (2007). Where Excludability Matters: Material Versus Intellectual Property in Academic Biomedical Research. *Research Policy*, 36(8):1184–1203.

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

Ziedonis, A. A. (2007). Real Options in Technology Licensing. *Management Science*, 53(10):1618–1633.

Table 1: Summary Statistics for Samples with 3 and 4–Author Publication Matches to Patents†

Samples		
Inventor Overlap	3+	4+
Publications (000)	1.7	0.6
Patents (000)	0.7	0.3
Publications with MTAs	261	79
Publications / Patent	2.4	1.8
Observations in the Life Sciences	49%	44%
Variable	**Mean**††	
Citations Per Year	11.4	16.2
	(26.1)	(36.3)
Average Impact Factor	8.7	11.0
	(8.4)	(9.8)
Publication Year	2000.7	2000.4
	(2.6)	(2.8)
Publication Age	3.2	3.3
	(2.6)	(2.7)
Age at MTA Issuance	2.4	2.6
	(2.7)	(2.6)
Age at Patent Issuance	3.5	3.5
	(2.0)	(2.0)

†Sample restricted to those publications with 7 years of citation data (only first 7 years of data included)
††Values in the parentheses are standard deviations.

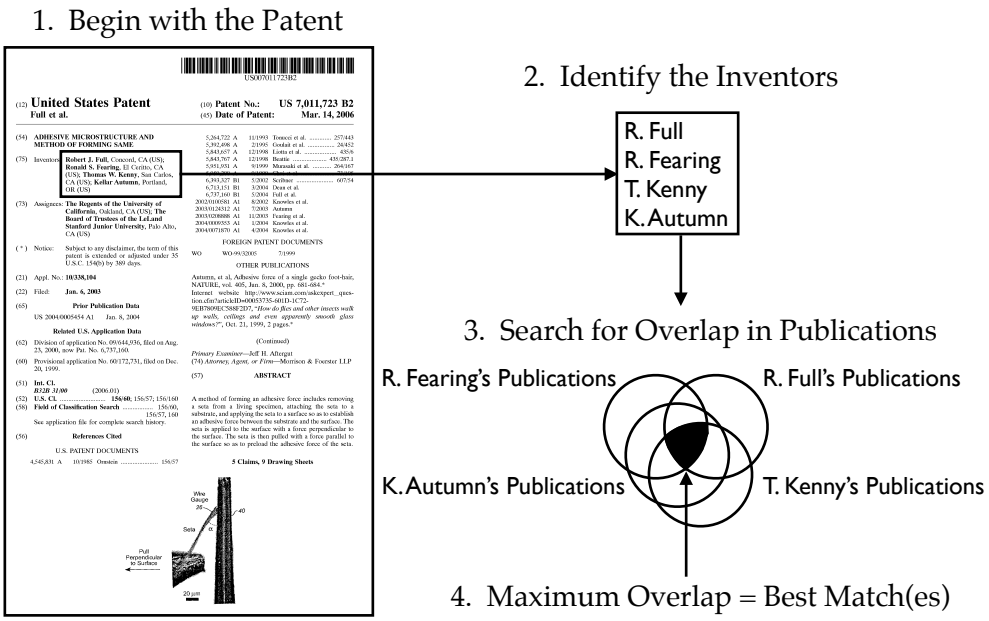


Figure 1: Inventor–Based Matching Example

Table 2: Covariate Balance for the Full Licensing Sample

Covariates	Full Licensing Sample			
	Mean Treated	Mean Control	*t*–Test	*KS* Test
Publication Age	2.0	2.0	-	-
Journal Subject†	4.3	4.3	-	-
Patent Issued	0.56	0.56	-	-
MTA Issued	0.04	0.04	-	-
Journal Impact Factor	7.3	7.1	-	-
Publication Year	2000.7	2000.5	*	-
Citations in t–1	6.6	6.1	-	***
Citations in t–2	4.3	4.2	-	-
Citation Slope from t–2 to t–1	2.3	1.9	***	***

Notes:
*** p<0.01 ** p<0.05 * p<0.10

† *Journal Subject* is a categorical variable, with each subject mapped to a random integer. Therefore the 4.3 listed has no literal meaning, but the equality between treatment and control, as well as well as the lack of any difference, reinforce the success of the exact matching.

Table 3: Covariate Balance for the Research Tools Sample

Covariates	Research Tools Sample			
	Mean Treated	Mean Control	*t*–Test	*KS* Test
Publication Age	1.9	1.9	-	-
Journal Subject	3.9	3.9	-	-
Patent Issued	0.5	0.5	-	-
MTA Issued	0.3	0.3	-	-
Journal Impact Factor	5.2	6.0	-	*
Publication Year	2001.4	2000.5	***	-
Citations in t–1	5.4	5.0	-	-
Citations in t–2	1.2	1.3	-	-
Citation Slope from t–2 to t–1	4.3	3.8	-	-

Note:
*** p<0.01 ** p<0.05 * p<0.10

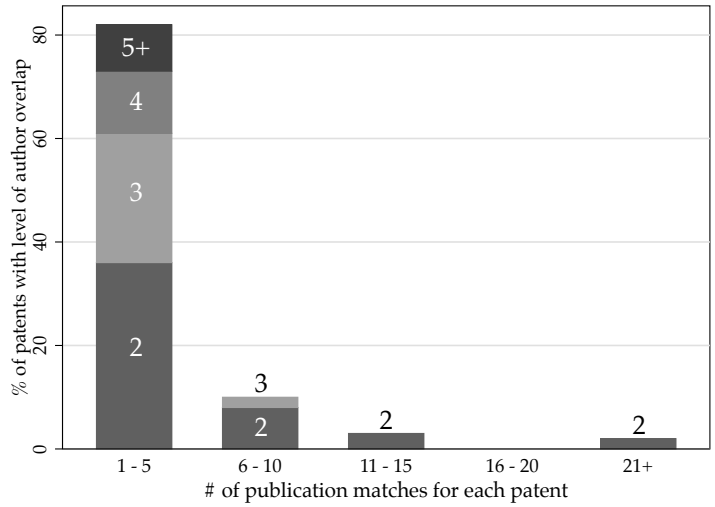


Figure 2: Distribution of Patent–Paper Pairs with Two or More Names Common to Both Matched Patent and Published Paper

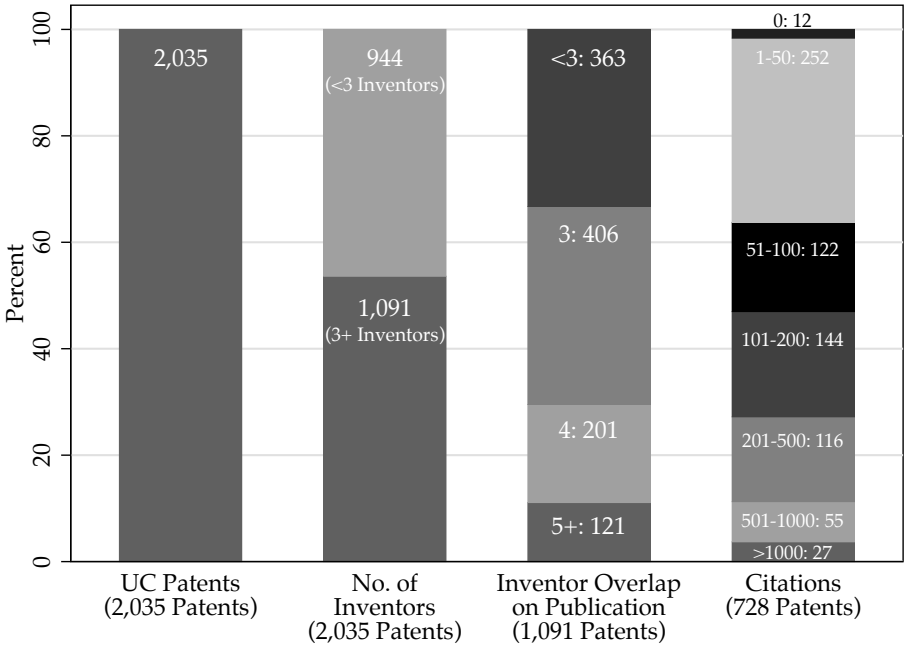


Figure 3: Sample Composition

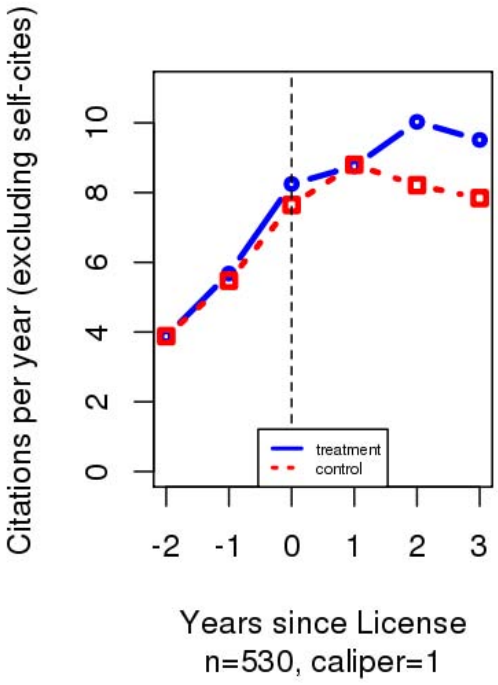
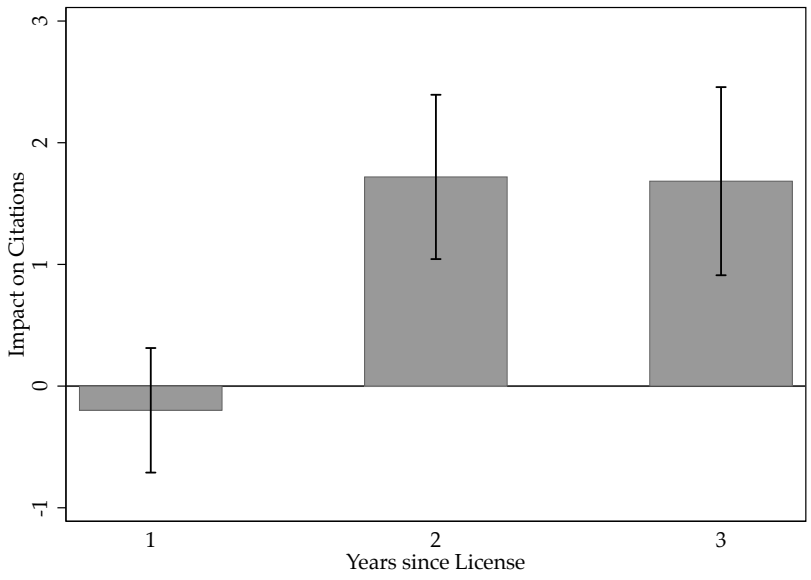


Figure 4: License Effect—Full Licensing Sample



(Error bars signify ± 1 standard error.)

Figure 5: Regression Results of License Effect–Full Licensing Sample

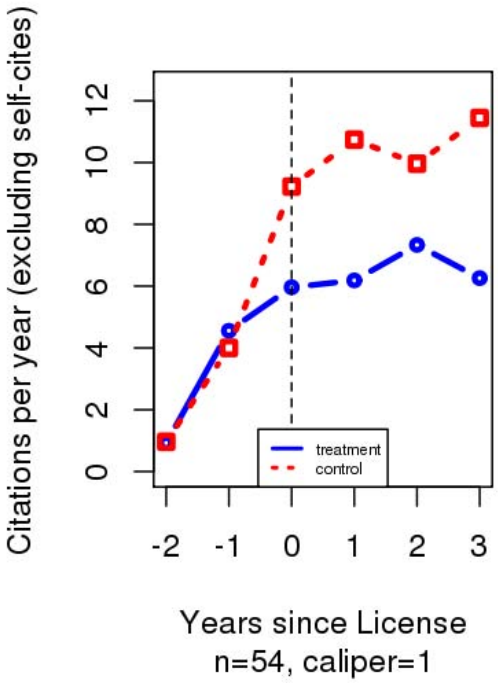
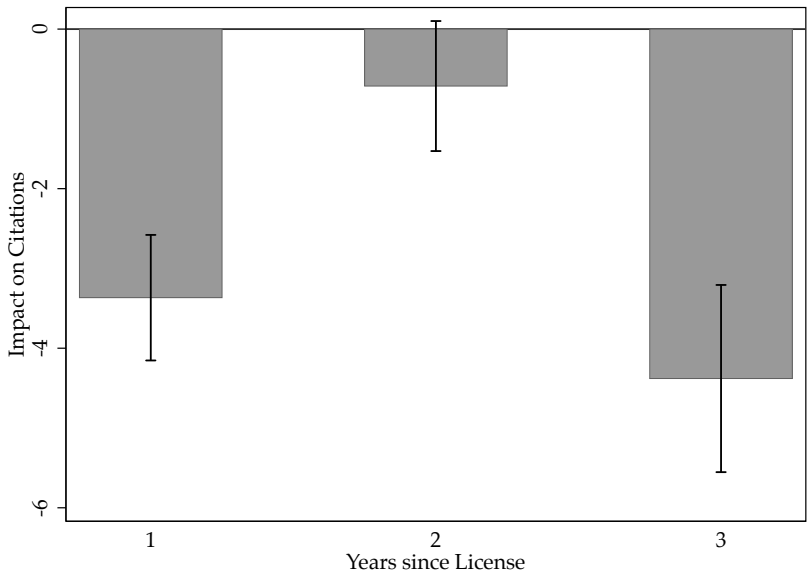


Figure 6: License Effect—Research Tools



(Error bars signify ± 1 standard error.)

Figure 7: Regression Results of License Effect—Research Tools