Hubs as lampposts: Academic location and firms' attention to science*

Michaël Bikard
London Business School
mbikard@london.edu

Matt Marx
Boston University Questrom School of Business
mattmarx@bu.edu

Abstract: Prior literature has described the importance of firms' attention to external knowledge for innovative performance, but it has focused exclusively on drivers of attention that are internal to the firm. We propose that external elements also shape firms' attention. Specifically, we show that firms' selective attention is drawn to academic discoveries that originate in "hubs" of commercial R&D in a particular field. Testing the impact of hubs on firms' attention is difficult because different papers may be more or less valuable for industry. We address this identification challenge by analyzing simultaneous discoveries where multiple researchers report the same finding in "twin" papers. Hubs may attract attention to local academic science by enabling informal interactions among academic scientists and industry inventors; indeed, we find that the hub effect is strongest in cities where many scientific conferences are hosted. Hubs' "lamppost effect" is moderated by institutional prestige and by formal connections between academia and industry.

JEL codes: O00, O32

*Authorship is alphabetical. For their valuable feedback, we thank Pierre Azoulay, Sharon Belenzon, Michael Ewens, Jeff Furman, Christopher Liu, Mark Schankerman, Eunhee Sohn, Scott Stern, Keyvan Vakili, and Ivanka Visnjic and participants at seminars at Boston University, Erasmus, the IFN, INSEAD, London School of Economics, the Stockholm School of Economics, UC Berkeley, and Wharton and at the Duke Strategy Conference, Georgia Tech Roundtable for Engineering Entrepreneurship Research, Israel Strategy Conference, and NBER Productivity Lunch. This work was supported by a Kauffman Junior Faculty Fellowship and by the Deloitte Institute of Innovation and Entrepreneurship at London Business School.

Introduction

Firms' ability to take advantage of external knowledge is a critical driver of performance. A large literature has highlighted how a firm's cognitive structures, resources, routines, and players affect its allocation of scarce attention and determine the types of knowledge that it might absorb (Simon 1947; March and Simon 1958; Cohen and Levinthal 1990; March 1991; Ocasio 1997). Those drivers, however, are all internal. It is possible that external elements also influence a firm's attention to external knowledge and therefore its capacity to absorb it. There might be sources of systematic variance in the visibility of external knowledge—independent of its actual relevance—that affect the menu of options a firm considers.

We focus on firms' attention to science and argue that firms that are developing new technologies systematically pay more attention to scientific discoveries from a hub—defined as a concentration of commercial R&D in a particular field—than to knowledge that emerges elsewhere. Not only does knowledge in those hubs disseminate across collocated firms "as if it were in the air" (Marshall 1895: 330), but we propose that hubs also draw the attention of noncollocated firms to local knowledge because they are areas of transit in which informal interactions across organizations take place regularly. Knowledge emerging in hubs might therefore be significantly more visible to firms, all else equal.

Testing this hypothesis empirically is not straightforward because knowledge emerging in hubs may differ fundamentally from knowledge originating elsewhere. Firms might pay more attention to scientific knowledge produced in hubs if that knowledge tends to be more relevant to them. However, there might be other reasons. To identify the effect of the location of knowledge on the attention firms pay to it, we exploit the occurrence of simultaneous scientific discoveries (Merton 1961).[1] When two or more scholars publish the same findings, they create "paper twins." We measure firms' attention to those papers by observing citations in their patents of sets of paper twins where one paper emerged from a hub and the other does not—with fixed effects for the set-of-paper-twins/patent dyad. Our main analysis focuses on 147 simultaneous discoveries disclosed through 316 academic publications and the citation or noncitation of those papers in 697 industry patents.

Our results suggest that firms pay more attention to academic publications when they emerge in hubs. In our data, academic papers of which at least one author is located in a hub are approximately 10% more likely to be cited in patents assigned to firms than papers with no authors located in a hub. A similar

[1] We focus on the natural sciences, but simultaneous discoveries occur in other fields. In economics, for example, the discovery of a competitive equilibrium in a market economy by McKenzie and by Arrow and Debreu in 1954 is one famous example (Weintraub 2011); other cases have also been reported (Stigler 1980; Niehans 1995).

effect is *not* observed for citations in patents assigned to academic institutions. It may be that academics already have a good view of the scientific literature and therefore do not rely on hubs' "lamppost effect." The effect of hubs on firms' attention is robust to a number of specifications and we find no evidence that it is driven by formal social networks. However, we do find evidence suggesting that it might be driven in part by informal interactions. In our data, the effect of hubs appears to be stronger for hubs that host numerous scientific conferences attended by both academics and firms.

This paper makes three contributions. First, while prior literature has highlighted the internal drivers of firms' attention to external knowledge (Ocasio 1997; Monteiro 2015; Piezunka and Dahlander 2015), our results show that external elements also matter. Second, we highlight one external element that systematically affects firms' attention to science other than the well-documented effect of distance (Jaffe, Trajtenberg, and Henderson 1993; Zucker, Darby, and Brewer 1998; Belenzon and Schankerman 2013): R&D hubs act as lampposts, increasing the degree to which local academic work is visible to industry. Third, our findings contribute to the literature on hubs (Jacobs 1970; Glaeser et al. 1992; Saxenian 1994; Padgett and McLean 2006) by highlighting that they foster the circulation of ideas not only among co-located organizations, but also beyond the hubs' borders.

1. Firms' attention to science

Whether deliberately or not, firms constantly select the knowledge that they will pay attention to and the knowledge that they will overlook (Simon 1947; March and Simon 1958; Cyert and March 1963). Simon (1947: 124) noted: "In the contemporary world all of us are surrounded by, even drowned in, a sea of information, only an infinitesimal part of which can be attended to. Although we may wish to have certain kinds of information that are not available (e.g., reliable forecasts), the critical scarce factor in decision-making is not information but attention. What we attend to, by plan or by chance, is a major determinant of our decisions." Given the importance of this process for firm performance, a vibrant literature has developed, over the past two decades, an attention-based view of the firm to examine how firms' rules, players, positions, and resources shape the allocation of attention (Ocasio 1997; Hansen and Haas 2001; Ocasio 2010).

Firms' attention to external knowledge affects innovation performance in a number of ways. Li and collaborators (2013) investigate the relationship between top management teams' attention and firm innovation in 61 public US high-technology firms, finding that innovation is driven both by the selection of what managers attend to and by the intensity of that attention. Monteiro (2015) examines managers' attention to external opportunities at a large telecommunication services provider and finds that managers at headquarters pay a disproportionate amount of attention to confirmatory and proven external knowledge, even when it is reported by a unit that was specifically established to overcome local search. Similarly, Piezunka and Dahlander (2015) undertake a study of 105,127 suggestions that were submitted to 922 organizations and find that organizations that solicit external ideas with the goal of gaining access to distant knowledge often end up filtering out that same knowledge.

It is noteworthy that the drivers of firms' attention highlighted in prior research are all internal. Yet, external elements might also influence the knowledge that the firm will attend to. After all, some opportunities are more salient than others and this variance might systematically affect the menu of options that firms will contemplate.

Discoveries published by academic scientists are an important type of external knowledge for innovative firms. Even though some academic science is patented or subject to license, the methods and findings of such research are communicated via public disclosure. Since the first scientific article was published in the *Journal des Sçavans* in Paris on January 5, 1665, over 50 million manuscripts have been published and the speed at which new contributions are added to the scientific commons is still increasing (Jinha 2010). The rapid advancement of public science constitutes a growing burden for researchers, who find they must lengthen their training, specialize, and collaborate in order to stay up to date (Jones 2009; Agrawal, Goldfarb, and Teodoridis 2016). The implications of this burden for firms, however, are less clear. Firms routinely invest in absorbing external scientific knowledge to increase their innovative performance (Cohen and Levinthal 1990). Considering the breadth of the scientific literature, the way they allocate their attention to it is likely to determine what they absorb, which will in turn shape the direction and efficiency of their invention efforts (Fleming and Sorenson 2004). Clearly, internal elements will affect their attention to the scientific literature; prior research indicates that firms are likely to pay more attention to academic discoveries that match their skills and experience. However, external elements might also matter.

This paper identifies one such external element: whether or not the academic authors of a particular paper are located in a hub. We propose that hubs of commercial R&D may act as "lampposts" that draw the attention of firms to local academic discoveries. Hub location might attract local firms' attention because of the sheer number of collocated firms (Jaffe, Trajtenberg, and Henderson 1993; Zucker, Darby, and Brewer 1998; Adams 2002; Zucker, Darby, and Armstrong 2002; Furman and MacGarvie 2007; Belenzon and Schankerman 2013). However, we propose that originating in a hub might also increase the attention a scientific publication receives from firms *outside* the hub. This proposition draws on a long tradition in the economic literature. Smith (1776) and later Mill (1848) proposed that proximity to ports, roads, and railways—i.e., location in hubs— enables access to more distant markets. Our argument does not focus on the role played by a hub's physical infrastructure, but rather on the role of informal interactions. R&D hubs are not only destinations to which corporate inventors are likely to travel, they are also settings in which informal interactions naturally take place between industry and academia. Since those informal interactions are a key channel through which public research influences industrial R&D (Cohen, Nelson, and Walsh 2002), one might naturally expect that local academic work will be more salient to firms.

Firms' attention to academic science can be measured by observing their citations of the scientific literature in their patents. Patent-to-paper citations are a relatively good measure of the flow of public science to firms (Roach and Cohen 2013) and an increasingly popular one (Azoulay, Graff Zivin, and Sampat 2012; Belenzon and Schankerman 2013). In the case paper twins, it makes no legal difference whether a patent cites one or the other or all of them. As we describe below, we find in 21 interviews that a firm's differential citation of papers from a set of paper twins reveals the differential attention it paid to those papers. Patent citations to paper twins thus provide a window onto firms' selective attention to the academic literature. We therefore hypothesize:

HYPOTHESIS. *Firms are more likely to refer in their patents to a publication from a hub than to a publication that discloses essentially the same discovery but is not from a hub.*

.

2. Data Construction

A. Empirical Approach

Identifying the impact of hub as lamppost attracting firms' attention to local knowledge is nontrivial because the emergence of knowledge in specific locations is not exogenous to the geographic distribution of firms. In particular, firms influence nearby academic research (Sohn 2014). For our purpose, this raises a well-known identification challenge. How can the empiricist "divine whether a particular citation would have taken place, if contrary to the fact, either the citing or the cited producer had been located elsewhere?" (Azoulay, Graff Zivin, and Sampat 2012, 13).

To address this challenge, we use simultaneous discoveries in the natural sciences. Because simultaneous discoveries constitute instances in which the same or very similar knowledge emerges in multiple locations, "paper twins" present a unique opportunity to unbundle producers from their products. Rather than creating a control sample of non-referenced publications or patents (Jaffe, Trajtenberg, and Henderson 1993), these paper twins allow us to measure firms' attention directly by examining patent references to the academic publications which make up each set of twins while accounting for the characteristics of the individual scientists and of their institution. The next three sections describe (1) the paper twins; (2) our measure of firms' attention to academic science; and (3) the way we identify whether a publication stems from a hub or not.

B. "Paper Twins"

This study is based on the first automatically and systematically collected dataset of simultaneous discoveries. Before describing the process by which such publications were found, we illustrate the nature of a simultaneous discovery with an example. The August 1998 issue of *Cell* contains two papers reporting the same scientific discovery, shown in Figure I. Both papers report the discovery of an important molecule involved in cell death or "apoptosis." The two teams found that after activation of the death receptors on the cell membrane, the death signal is carried to the mitochondria by a cytosolic protein called BID. Confirming that these two papers truly report the same scientific discovery, an August 21 2000 article in *The Scientist* notes that "[t]hese two *Cell* papers outline two independent identifications of a critical missing link in [the apoptosis] signaling pathway" (Halim 2000). Frequently in the case of simultaneous discoveries authors send their manuscripts to the same journal, sometimes leading to back-to-back publications[2] (in this case: pages 481-490 and 491-501; in our dataset, 46% of the simultaneous discoveries are published back-to-back in the same issue of the same journal).

Figure I about here

To detect simultaneous discoveries, an algorithm was built that identified frequently co-cited pairs of papers and then scrolled through the scientific literature to spot instances in which two papers are consistently cited *in the same parenthesis*, or adjacently. The method is detailed in a companion paper (Bikard 2012), but we present a summary here for convenience.

Sociologists of science have found that citations provide a window into the scientific community's allocation of credit (Cozzens 1989). While two publications that are consistently cited in the same literature might be complementary rather than overlapping, the fact that co-citations are consistently adjacent suggests that the credit for a discovery is being shared. The algorithm that was used therefore considers pairs of scientific publications that are consistently cited together—i.e., in the same parenthesis, or adjacently. The algorithm proceeds as follows. A dataset consisting of 42,106 scientific articles published between 2000 and 2010 was built from the 15 non-review scientific journals having the highest impact factor in 2009. Three-quarters of a million unique references from these articles were grouped into pairs that (a) were co-cited at least once, (b) were written no more than a calendar year apart, (c) have

[2] Editors sometimes publish manuscripts back-to-back, recognizing a tie in the race for priority and allowing both teams to receive equal credit for their work. Well-known examples of back-to-back publications include that of evolution by natural selection by Darwin and Wallace in the *Journal of the Proceedings of the Linnean Society of London* published on 20 August 1858 and the discovery by Richter and Ting of the J/ψ meson published in *Physical Review Letters* on 2 December 1974. While simultaneous discoveries appear often (but not always) back-to-back in scientific journals, not all back-to-back publications correspond to simultaneous discoveries (Drahl 2014).

no overlapping authors, (d) in which at least 5 citations for each reference are observed. A Jaccard co-citation coefficient was then calculated for all pairs meeting these criteria; pairs with a Jaccard coefficient >50% were retained. Finally, those pairs for which 100% of the co-citations took place in the same parenthesis or adjacently were retained. These 720 pairings of 1,246 papers disclose 578 unique discoveries since there are instances of discoveries involving three or more teams.

From this set, we discard 80 papers published by firms or whose only twin is a firm-published paper as our aim is to understand which academic papers firms focus on. We also drop 804 publications from sets of paper twins where none of the papers received any references from patents assigned to firms. This could happen if none of the papers were referenced by any patent, if they were referenced only by academic patents (which we will later use in a placebo test), or if the citation was a self-citation. Finally, we removed 17 publications where any patent referencing one paper from a set of paper twins references all of the papers from that set (and thus does not provide variation on our dependent variable; however, our results are robust to using linear probability models, which include these). Excluding those leaves 345 publications.

For each publication, we collect its geographic origin, journal, and whether the discovery was itself patented. (Whether the focal paper reporting a simultaneous discovery was patented by its authors is an essential control as doing so forms a "patent paper pair" (Murray 2002).) To account for the author heterogeneity, we collect the corresponding author's stock of patents and papers at the time of publication. Similarly, we capture the institution's stock of patents (past five years). As a measure of the institution's prestige in the particular field of that paper, we count the number of papers ever published by that institution in the top 15 scientific journals relevant to that paper, as calculated from 1,858 Medical Subject Heading (MeSH) keywords since 1955. For each article, we gathered the ISI classification most frequently associated with each of its MESH keywords and produced an ordering based on the frequency of ISI classifications for those keywords.[3] Summary statistics for all 1,166 academic twin papers are in Table I, segmented by whether none of the twins were cited by commercial patents (804), the twins were always cited jointly (17), sometimes cited jointly (138), or cited but never jointly (207).

Table I about here

Table II provides a breakdown of the most frequent cities and institutions among the academic publications in our analyses, for which one but not all papers of a set of paper twins were referenced. As

[3] The ISI keywords for our journals are Biochemistry & Molecular Biology, Biotechnology & Applied Microbiology, Cell & Tissue Engineering, Cell Biology, Physical Chemistry, Genetics & Heredity Immunology, Materials Science, General & Internal Medicine, Research & Experimental Medicine, Nanotechnology, Oncology, Applied Physics, Condensed Matter Physics, and Interdisciplinary.

is visible in Panel A, many of our papers stem from high-status institutions, including nearly 5% from Harvard University alone. This underscores both the importance of accounting for the prestige of the institution (since industry inventors may be more likely to be exposed to such discoveries). A similar concern also applies in Panel B, which shows that nearly 20% of all twin papers are published in Boston, New York, and San Diego.

Table II about here

C. Measuring Firms' Attention to Academic Science

Determining whether firms pay attention to academic science is challenging. A firm is a collection of individuals, and reliably tracking whether each employee paid attention to every relevant piece of scientific knowledge would involve a daunting amount of fieldwork. However, among some of the most innovative employees of the firm—namely, patenting inventors—one proxy for their attention is the works they cite in their patents. One possibility is to look at patents cited by patents assigned to firms, originally pioneered by Jaffe, Trajtenberg, and Henderson (1993), but these have substantial limitations. First, patent citations are often added by patent attorney and patent examiners (Alcácer and Gittelman 2006; Alcácer, Gittelman, and Sampat 2009) and can be used strategically (Lampe 2012), obfuscating whether the inventor was actually aware of or built upon the cited work. Second, each patent is unique, making interpretation of non-citation difficult since each applicant makes decisions about the relevance of a given article based on the particularities of the patent in question. Concerns regarding the definition of a control group of non-citing patents has led to major debates in the literature (Thompson and Fox-Kean 2005; Henderson, Jaffe, and Trajtenberg 2005). Third, and perhaps most importantly for the present study, only a tiny fraction of academic knowledge is patented. Instead, academic institutions primarily disclose knowledge through scientific publications (Agrawal and Henderson 2002; Belenzon and Schankerman 2013; Roach and Cohen 2013).

Instead using citations from patents to academic papers (Belenzon and Schankerman 2013; Azoulay, Graff Zivin, and Sampat 2012) helps overcome these challenges. Roach and Cohen emphasize that "citations to nonpatent references, such as scientific journal articles, correspond more closely to managers' reports of the use of public research than do the more commonly employed citations to patent references" (Roach and Cohen 2013, 505) and thus should afford a better sense of where the firm allocates its selective attention. In addition, the simultaneous discoveries in our sample are overwhelmingly from the life sciences (to which most of the highest-impact-factor scientific journals belong), which is advantageous because in the life sciences the use of publications and patents by firms is widespread (Branstetter and Ogura 2005).

Another advantage of using patent-to-paper references stems from the fact that although every patent by definition represents a unique invention; the patent system does not recognize "ties" in the race for priority; thus simultaneous or independent inventions are debated by legal scholars (Vermont 2006; Lemley 2007). But the same is not true for scientific publications: when two researchers make the same discovery and send it for publication at around the same time, multiple papers are frequently published disclosing very similar knowledge (e.g., Cozzens 1989). Hence, it may be that an inventor is not aware that a discovery was reported in two or more academic publications—and that all of those publications could in principle be referenced in the patent as prior art. In practice, the U.S. Patent and Trademark Office imposes no duty on the inventor to reference *every* paper disclosing the same simultaneous discovery. According to USPTO Rule 56 (37 CFR 1.56): "information is material to patentability when it is not cumulative to information already of record or being made of record in the application." In other words, if multiple papers disclose the same knowledge, referring to just one is sufficient. Even if individual inventors are not aware of Rule 56, its existence entails that patent lawyers need not pressure inventors to be exhaustive in coming up with a list of patent-to-paper references in the case of paper twins.

Despite the growing literature using patent references to the academic literature as a measure of knowledge dissemination (Narin, Hamilton, and Olivastro 1997; Agrawal and Henderson 2002; Azoulay, Graff Zivin, and Sampat 2012; Roach and Cohen 2013), the actual mechanism underlying such references remains little understood (Callaert et al. 2006). To our knowledge, the actual process through which those references emerge has not yet been systematically studied. To uncover the "data generation process" for our data specifically, we interviewed 21 inventors in July-October 2016. We identified the 50 individuals that were listed as inventors in a large number of patents while avoiding instances in which two individuals are repeatedly co-inventors on the same set of patents. 21 of the 50 inventors accepted to answer our questions over the phone or in person. They are researchers occupying senior R&D positions at a variety of pharmaceutical or biotechnology firms in the US and in Europe. Those interviews do not constitute a comprehensive (or representative) survey of inventors' citation practices, but they do provide a window onto an important process that has received to date very little scrutiny. In particular, three themes emerged from those interviews.

First, patent references to the academic literature seem to be generally included by the inventors, not by the attorney or the examiner (Carpenter and Narin 1983). While our interviewees generally outsourced the patent-to-patent citations to their attorney, the same is not true of their patent-to-paper citations. About half of our inventor interviewees reported that they produced 80-100% of the academic references in their patents, and nearly a third said that they produced over 60% those references. The few inventors that reported less involvement mentioned that their attorneys used templates when writing patents and that those templates typically already involve a number of scientific references.

Second, we found that most inventors that we interviewed do not try to be exhaustive when it comes to citing the academic literature in their patents. In this sense, the process is very different from that of citations in academic publications. Only three of our 21 interviewees believed that exhaustiveness is important when it comes to academic references in their patents. Said one inventor, "*scientists are not as careful for patents as they are for their own publications (…)*" Another recalls: "*You know, it is sort of an abbreviated list of citations that establish a certain sense of what's known in the art but it's not necessarily exhaustive or for that matter even fair in terms of giving credit*."

Third, our interviewees indicated that the main reason why one paper gets referenced and not another is inventor's degree of awareness of or familiarity with the paper. We asked "why did you cite this paper and not this other one in this patent?" Eighteen of our interviewees mentioned that they were less familiar with the paper that they did not cite, and 14 mentioned that they might simply not have known about the other paper. As an example, one inventor answered: "*The more familiar paper comes to mind first, plus the citation makes the point. Done! I think it is convenient to say that we know about all the other papers, but there might be examples where in fact we don't know. I can't say that it is always true that we were aware of everything.*" In some cases, the inventors did not know the twin that they did not cite. However, our interviews suggest that systematic differences in references are likely to be primarily driven by the degree of attention paid to those papers.

Perhaps surprisingly, given the prevalence of "strategic" patent-to-patent citations (Lampe 2012), only a few of our interviewees admitted to such behavior. Moreover, those who noted the possibility of strategic citations made opposite predictions about its direction. One interviewee highlighted that he would be more likely to reference a competitor so as to demarcate his work, while two other interviewees made the exact opposite argument, that some inventors might avoid citing their competitors. Although this is a very small sample, strategic citation does not seem a key factor for patent-to-paper citations.

In sum, in exploring the data generation process in our setting, we conducted what is to our knowledge the first attempt to provide systematic descriptive information about the process through which inventors in firms cite the academic literature in their patents. The number of interviews is small (21 with a 42% response rate) and these results should therefore be interpreted carefully. However, our results on patent-to-paper citations are strikingly different from what we know about patent-to-patent citations and to paper-to-paper citations. This is noteworthy because all three types of citations have been used in prior literature as measures of knowledge flow. Our interviews highlight that patent-to-paper citations in our setting may reflect the relative amount of attention that inventors paid to different scientific publications.

Unlike patent-to-patent citations, which are presented in a formatted table in the header of the patent record, references to scientific publications are not straightforward to analyze. As is visible in

Figure II, patent-to-paper citations are listed as unstructured strings. One might match on title and journal name, but we found that names were abbreviated so often and in many different forms as to rule out such an approach. (For instance, *Nature Genetics* is abbreviated as Nat Genetics, Nat Genet.; Nature Genets, etc.) Titles are often omitted altogether. Instead, we used four simpler criteria: 1) surname of the first author 2) year of publication 3) volume of journal 4) starting page number. The resulting tuple is likely unique to the paper; nonetheless, we inspected by hand the resulting matches. We use patent data from Dataverse (G.-C. Li et al. 2014). We checked patents citing twin papers by hand to categorize them as academic or commercial (i.e., firm-assigned) patents.

After locating all patent-to-paper references, we exclude two types of self-references (results are robust to including self-references). First, if the surname and first initial of any author on the paper matches any inventor on the patent, we remove the paper-patent dyad from consideration. (References from within the same organization, typically excluded from patent-citation studies, are not of concern because the patents are from firms while the papers are from academic institutions.) Second, we reviewed the acknowledgments section of each paper and then excluded references where the patent assignee was acknowledged as a sponsor of that research. This exercise yields our dependent variable *REFERENCED*.

Figure II about here

D. Determining whether a paper is in a hub

Our analysis turns on whether a given paper is in a "hub" – i.e., a geographic concentration of industrial R&D relevant to the paper. We do this by detecting whether any of the authors on the paper is within commuting distance (50 miles) of geographic patent concentrations in relevant technological subclasses. Details of this process are described below.

We start by collecting the technological subclassifications from all patents, whether industry or academic, that contain references to one of the papers of a set of twins in order to have the most complete possible representation of USPTO patent subclasses that are applicable to the discovery. Patents referencing papers that report the simultaneous discoveries are from 712 technological subclasses. For each of those subclasses, we collect all firm-assigned patents belonging to that subclass, whether or not they reference any of the twin papers. We find a total of 101,441 industry patents that were categorized by the USPTO into one of those 712 technology subclasses and note the location of each patent.

For each of the 712 technology subclasses related to the simultaneous discoveries, we count the total number of industry patents for each half-decade and also the number in each city during that same time window. For each location, we divide the latter figure by the former to obtain the percentage of patenting activity in that subclass occurring there during that five-year period. A location is classified as having a geographic concentration of patenting for a given subclass if it satisfies two criteria. First, more than 5% of patents in that technology subclass must be in that location. Second, because this threshold can easily be exceeded in technology subclasses with few patents (e.g., in a subclass with only 20 patents, *every* location has at least 5% of patenting), a location must have at least five patents in that subclass to qualify.

This exercise yields a list of cities where patenting is geographically concentrated for the 712 technology subclasses relevant to our simultaneous discoveries. (Some subclasses are geographically dispersed and thus do not have any such entries.) We then form a "hub" with that city in the center and sweeping a circle with a radius of 50 miles. We chose 50 miles as an approximation of reasonable commuting distance. The results presented below are not insensitive to this 50-mile assumption; although restricting to shorter commuting distances yields similar results, loosening the restriction by allowing commuting distances of 90 miles or greater fails to yield statistically significant estimates.

As a final step, we check whether any of the authors on a given paper resides in a hub corresponding to the relevant subclasses for that paper, as determined above. Each author's location is geocoded from their affiliation on the paper.(Calculating whether the paper is in a hub using only the location of the principal investigator yields similar results.) Applying this definition, 23.7% of the publications in our paper-patent dyads are in a hub. In many cases, all the papers of a set of paper twins are located outside a hub. Hubs for the 712 subclasses are shown in Panel A of Figure III; Panel B provides a U.S.-only zoom.

Figure III about here

To illustrate the concept of being located in a hub, we return to our simultaneous discovery from Figure I: two papers in the August 1998 issue of *Cell*, one at Harvard Medical School in Boston, MA and another at UT Southwestern Medical Center in Dallas, TX. (For both of these papers, all authors have the same affiliation.) In determining whether either of these research teams was in a hub, we first note that 19 patents list one of these papers as a scientific reference. We then define the scope of relevant R&D by obtaining the USPTO technological subclasses for these patents. A few of the patents have the same subclass, yielding 17 unique subclasses. The next step is to locate geographic patenting concentrations in these technological areas. We find 3,858 firm-owned patents that were assigned to these subclasses during 1995-1999. The patenting concentrations for the 17 subclasses include Milan, Italy; La Jolla, Santa Clara, and Solana Beach, California; Canton, Massachusetts; Silver Spring, Maryland; Berkeley Heights, New Jersey, and Bainbridge Island, Washington. We then check whether either institution is within commuting distance of those concentrations. Boston is within 50 miles of Canton, Massachusetts, but Dallas is far from any of the cities listed so only the Boston-based paper is "in a hub."

F. Empirical Setup

Our analysis leverages the simultaneous-discovery nature of our data because a patent that references one paper is presumably at a similar risk of referencing any of its twins. An observation is a dyad of a published paper reporting a simultaneous discovery and a patent at risk of referencing the paper. To obtain a dataset that includes not only realized patent-to-paper citations but also unrealized citations to the twins of a focal paper, we pair each industry patent[4] that references a paper with the other "twin" papers that disclose the same simultaneous discovery. That is, given a set of paper twins where one of the papers is referenced by a later patent, we also create an observation for that same patent together with the twin paper that was not referenced but could have been. This process yields 1,671 paper-patent dyads.

For each paper-patent dyad representing a (potential) scientific reference, we account for both temporal and spatial separation between the paper and the patent. Given that our key explanatory variable reports whether a paper is located in a hub, the distance between paper and potentially-referencing patent is perhaps our most important control. One might worry that papers in hubs are simply within shorter distances of potentially-citing patents and thus may be more likely to be referenced by those patents for reasons previously established in the literature on distance and knowledge diffusion. We control for distance in a linear fashion with the logged count of miles between the paper and potentially-referencing patent. Also, twin papers are usually but not always published in the same calendar year, so we control for the lag between the publication of the twin and the potentially-referencing patent. Summary statistics are in Table III.

Table III about here

We specify a conditional logit model with fixed effects for the simultaneous discovery and a focal patent that references one but not all papers of the set of paper twins reporting that discovery. Thus, for a given patent that is arguably at equal risk of referencing any of the paper twins, our analysis reveals the factors associated with a particular twin being referenced. The regression equation is given as

$$REFERENCED_{ijk} = f(\varepsilon_{ijk};\ \alpha_0 + \alpha_1 IN\_HUB_i + \alpha_{2} \circ \bar{X}_{ik} + \gamma_{jk})$$

where *j* represents the simultaneous discovery, *i* represents the paper reporting the simultaneous discovery, and *k* represents the potentially-referencing patent. IN_HUB_i is the main explanatory variable and is defined at the paper-patent dyad level. γ_{jk} is a simultaneous-discovery/patent fixed effect, which allows us to be unconcerned with characteristics of the patent (such as the assignee) other than in relation

---

[4] Twin papers are cited by three general types of patents: those assigned to firms, those assigned to academic institutions, and unassigned patents which remain the property of the inventor(s). We reviewed every patent that cited any of the twin papers by hand in order to determine whether it was firm-assigned, university-assigned, or unassigned. Only firm-assigned patents are included in our principal analysis, although we use university-assigned patents as a placebo test. Unassigned patents do not enter into our estimations.

to one twin paper vs. another.[5] Finally, X_{ik} is a vector of covariates including the geographic distance between the focal paper and the potentially-referencing patent. Standard errors are clustered at the level of the set of paper twins.

3. Results

A. Basic Results

We begin our analysis in Table IV. In column (1) distance is negatively associated with the likelihood of a paper being referenced by a patent. This is broadly consistent with Belenzon and Schankerman (2013) finding that the likelihood of a firm citing an academic paper is decreasing in distance and helps to allay concerns that this set of twin papers might exhibit unusual characteristics.

Table IV about here

Column (2) of Table IV tests our hypothesis that hubs attract firms attention to locally produced academic research. In practice, we include an indicator that an author of a paper was located in a hub of relevant commercial R&D. The estimated coefficient on distance diminishes somewhat in magnitude compared to column (1). Paper twins located in hubs are 10.1% more likely to be referenced than their twin(s) located outside those hubs, with statistical significance on the estimated coefficient at $p<.023$. Thus it appears that the deleterious effect of being separated from a potential industry consumer of academic knowledge can be ameliorated if the academic scientist is in a hub.

One concern is that our results could be driven by particular cities, especially those endowed with specific infrastructures (say, an airport) or those that host large concentrations of either academic scientists or industry inventors. In column (3) we introduce city fixed effects for the corresponding author's location. The resulting estimate of the coefficient on hubs has somewhat stronger magnitude and similar statistical significance ($p<.045$). Moreover, the introduction of city fixed effects reduces both the magnitude and statistical significance of the distance covariate to almost nil.

Column (4) adopts an alternative definition of a "hub." Instead of defining a publication as stemming from a hub if any of its authors are within 50 miles from a patenting hotspot, we simply count

[5] Of course, characteristics of the inventor, assignee, or patent itself may influence the construction of patent-to-paper citations. In unreported results, we replace simultaneous-discovery/patent fixed effects with fixed effects only for the simultaneous discovery. Although a less strict specification, this allows estimation of patent-level covariates. We cannot establish a connection between patent-to-paper citations and 1) number of inventors on the patent 2) number of prior patents per inventor 3) number of prior patents awarded to the assignee 4) whether one of the inventors is in a hub relevant to the focal paper.

the (logged) number of patents in the relevant subclasses within 50 miles of the corresponding author's location. Although this measure ignores the fraction of patenting in the subclass and thus does not compare the concentration of relevant R&D activity vs. other locations, it does make use of more information regarding the magnitude of patenting activity. The estimate of the coefficient for the count of patents within 50 miles is positive with statistical significance at the $p<.001$ level.

In column (5) we perform a placebo test that academic papers published in a hub are cited not just by firms but also by fellow academics. The analyses in Table IV until this point measure references to academic papers from patents assigned to firms in order to measure firms' attention to academic research. If hubs draw attention to local academic papers only from firms, the geography of industrial R&D should not affect citation rates *from within academia*. For column (5) we construct dyads of paper twins and university-assigned patents, using a setup similar to that described in Appendix III. The estimated coefficient on being in a hub is not statistically significant at conventional levels ($p<.133$).

B. Mechanisms

The foregoing analysis establishes a connection between the location of academic scientists in hubs and the subsequent likelihood of their articles being cited by industry patents but shines little light on the mechanisms underlying this connection. Next we explore possible mechanisms, including within-twin heterogeneity, formal networks, and informal interactions that arise from co-location.

Regarding heterogeneity, although we refer to the simultaneous discoveries as "paper twins" they are not entirely identical. Teams working on the same discovery might use slightly different approaches and surely describe their findings in different ways, even when the underlying discovery is the same. We assembled an expert panel to evaluate differences between the twin papers, hiring ten postdoctoral researchers in life sciences and applied physics from Imperial College, Cambridge University, University College London, and King's College. We focused their analysis on the 86 pairs of twins where one twin is in a hub and the other is not.

The expert-panel investigation was organized in two phases. In phase one, pairs of postdoctoral researchers were asked to jointly evaluate several sets of paper twins and highlight the dimensions along which they differed most. Our panel reported four dimensions along which they could distinguish paper twins: 1) level of details 2) strength of the claims 3) clarity of exposition 4) clinical relevance. In phase two, we asked the expert panel to evaluate the sets independently and to contrast the papers of the same set along all four dimensions. We then translated the responses into variables, creating dummy variables for each dimension that takes the value 1 if at least one of the postdoctoral researchers believed that one paper was superior to its twin along the specific dimension and 0 otherwise.

Table V presents results from our expert panel. The first two columns contain the means of those variables for hub and non-hub papers respectively. The third column evaluates whether the mean difference is statistically significant. Heterogeneity was more apparent in some dimensions than in others. For example, our experts were able to tell that one paper was more detailed than the other or that it was more clearly written in about one-third of the pairs. In comparison, they saw differences in the papers' clinical orientation in barely 7% of the pairs. We found that hub-based papers appear to make broader claims than their non-hub twin in 22% of the cases whereas the opposite was true in only 6.7% of the cases ($p=0.042$). On the other hand, we did not find any statistically significant difference in the amount of details, the clarity, or the clinical orientation of hub- and non-hub-papers. The only source of within-twin heterogeneity that clearly differs between papers in hubs vs. those not in hubs is the breadth of claims.

Table V about here

However, in column (1) of Table VI the breadth of claims is not strongly correlated with the likelihood of being cited ($p<0.162$). In fact, the only source of heterogeneity correlated with citation likelihood is clinical orientation (marginal effect = 10.3%; $p<.050$). As noted in Table V, however, clinical orientation does not differ significantly among hub vs. non-hub papers. Moreover, unreported regressions interacting the hub variable with each measure of heterogeneity fail to achieve statistically significant estimated coefficients on any of the interaction terms. Hence, it does not appear that heterogeneity among papers belonging to the same sets of twins explains the hub effect.

Table VI about here

Another possible explanation for the hub effect might be found in social networks. Singh (2005) shows that citations from one firm's patent to another firm's patent are strongly explained by patterns of co-inventorship, but it is unclear whether similar networks would facilitate the flow of knowledge *between* communities such as from academia to industry. We construct networks between academic and industry as overlap between authors on twin papers and inventors on patents citing them, including second-degree connections (i.e., any of the coauthors of any author on each paper overlapping with any of the co-inventors of any inventor on the patent). Column (2) of Table VI includes our measure of cross-community network overlap. The estimated coefficient is not statistically significant at conventional levels ($p<0.926$). Thus it does not appear that the hub effect is explained by networks.

Our failure to find that academic-industry networks explain why firms cite particular academic papers, although initially puzzling, may be explained in part by an observation from Dr. Michael Rogers, Research Scientist at Boston Children's Hospital and Professor of Surgery at Harvard Medical School: *"The industry network and the academic networks connect distally within themselves but only connect to each other locally."* In other words, it may be that hubs fill the gap in formal networks between academia and industry by facilitating interactions between academic and industry scientists. Such interactions may arise when scientists from the two communities attend seminars or bump into each other at social engagements. Of course, informal interactions are difficult to observe compared to the networks visible in co-authorship and co-invention. However, Dahl and Pedersen (2004)'s extensive survey work has documented that workers in the northern Danish telecom industry shared knowledge with those at other firms where no formal alliances or collaboration existed. In the remaining columns of Table VI we examine one observable proxy for informal interactions among commercial and academic scientists as an example of how these might arise in hubs.

Our proxy for informal interactions is the number of conferences held in a specific scientific field. The assumption is that when academic and industrial scientists attend the same conference, knowledge will be more likely to flow between these communities. Field experiments have shown that informal interactions such as at conferences can spur creativity and innovation (Boudreau et al. 2017). We combined the locations and topics of all Gordon Research Conferences—which are attended by scientists from academia, government, and industry—since 1970 with data from allconferences.com in the same set of scientific fields used to classify organizational prestige. In all, we found 2,383 academic conferences. Importantly, conferences were not exclusively or even primarily located in hubs (781, vs. 1602 in non-hub locations). Popular conference destinations include Bar Harbor, Maine and Santa Fe, New Mexico, which were not hubs for any scientific field among our twin papers. Conferences were geocoded and categorized according to sub-field and then matched to each of the papers if within 50 miles of the principal investigator's location.

In column (3) and (4) of Table VI we include the number of conferences relevant to the focal paper. The number of nearby conferences in a relevant field is somewhat predictive of the academic paper being cited ($p<0.078$), as seen in column (3), but the hub effect persists. Moreover, in column (4) a median split of the hub definition on the number of conferences reveals that the hub effect obtains principally among hubs with a higher number of conferences. Relative to non-hub locations, hubs with an above-average number of conferences have a marginal effect on citation of 10.1%, with statistical significance at $p<0.026$. By comparison, the estimate of hubs with a below-average number of conferences is imprecisely estimated ($p<0.613$). In unreported results, the distinction between hubs with more vs. fewer conferences is preserved when splitting by 75th, 90th, or 95th percentile. (Note that all results include a control for the population of the corresponding author's city.) We emphasize that these results should not be interpreted to mean that conferences are the only mechanism by which hubs promote awareness of academic knowledge among firms, rather than they represent an observable instance of the sort of informal interactions between academic scientists and industry inventors that may arise in hubs.

C. Heterogeneity across academic institutions

Given our hypothesis that locating in hubs will facilitate awareness by a larger set of industry inventors, we expect that our results will be attenuated by characteristics of the academic institutions that might attract firms' attention to the focal paper.

First, formal relationships between academia and industry may substitute for the informal interactions that may arise given location in a hub. (Given that this data is collected from the Association of University Technology Managers, our analysis is necessarily limited to North American institutions.) In column (1) of Table VII, we introduce a set of indicators interacting whether the paper is in an R&D hub with four levels of commercial funding of R&D at the institution: (1) no funding ("none"); (2) funding totaling less than \$5MM ("little"); (3) more than \$5MM but less than \$22MM, i.e. the 75th percentile ("more"); (4) more than \$22M ("most"). The interactions of hub with "none" and "little" are significant at $p<.028$ and $p<.030$, respectively, whereas the estimated coefficient on the interaction of hub with "more" funding is $p<.058$ and with "most" is $p<.460$. We graph the coefficients in Panel A of Figure IV. It appears that academic institutions that otherwise lack formal connection to industry benefit most from being in a hub.

Table VII and Figure IV about here

Next, we consider whether institutional reputation or prestige might generate exposure for papers located outside of hubs by compensating for those papers' lower visibility among commercial inventors. We interact the in-hub indicator with indicators for a median split of institutional prestige in column (2) of Table VII and rely again on the graphed coefficients for inference in Panel B of Figure III. The estimated coefficient on hubs interacted with below-median prestige is significant at $p<0.001$ whereas with above-median prestige fails to achieve statistical significance ($p<.379$). It appears that institutions with below-average prestige in the field of the focal paper benefit most from being in a hub.

Finally, we examine the interplay between hub location and the distance between the paper and the potentially-referencing patent. Following Singh and Marx (2013), in column (3) of Table VII we switch from a linear distance covariate to a non-parametric series of dummies in order to capture the nuances of localization. We add interaction terms for all distance dummies. The pattern revealed in the plot of linear-probability coefficients in Panel C of Figure III suggests a three-tiered effect of separation from relevant R&D hubs. For paper-patent dyads very close to each other (i.e., less than 20 miles), being in an R&D hub appears not to have an effect. This is visible as the estimated coefficient on the interaction of hub and no more than a 20-mile separation is significant at only $p<.623$. It appears that when an academic scientist and an industry inventor are already co-located, the latter is likely to pay attention to the discovery regardless of whether the former is in a hub.

For paper-patent separation more than 20 miles but less than 2000 miles, it appears that hubs help attract firms'attention. The estimated coefficient on hub interacted with separation between 20 and 2000 miles is positive and significant at $p<0.001$. By contrast, when the separation between the focal paper and focal patent is very large, hubs appear to be less efficacious. This in reflected statistical significance of p<.766 on the estimated coefficient of hubs interacted with a separation of more than 2000 miles between paper and patent. One might be concerned that this reflects the influence of bicoastal networks between say, San Diego and Boston. However, less than half of paper-patent dyads separated by more than 2500 miles are within the U.S. (39.9%), while all but one of the non-U.S. dyads with similarly large separation are on different continents.

In sum, Table VII and Figure III suggest that academic discoveries in hubs attract firms' attention mostly 1) when the author's institution lacks formal connections with industry, 2) when the institution has low prestige in that paper's scientific subfield, and 3) when the paper and potentially-citing patent are not already collocated. However, it appears that there are limits to the ability of hubs to draw firms' attention to academic research across extremely long distances.

4. Discussion

We interpret our results cautiously. The fact that an academic paper wasn't cited in a patent does not mean that the inventor was unaware of it. Hence, the economic significance of the variance in attention paid to knowledge from inside and outside R&D hubs is difficult to assess. In addition, our sample of twin papers is relatively small and largely concentrated in the life sciences (albeit not by design), limiting the generalizability of our results to other industries and to different types of external knowledge. Nevertheless, our setting provides a unique opportunity to study the impact of the location of external knowledge on firms' attention by exploiting the occurrence of simultaneous discoveries.

We find that academic articles of which at least one author is in a hub of industrial R&D are 10% more likely to be cited in firm-assigned patents than articles with no authors located in a hub. Results are robust to the addition of city fixed effects and to defining hubs according either to the relative concentration of activity or to absolute levels of activity. Moreover, location in a hub only draws the attention of inventors based in firms and not those based in academia. Hubs' "lamppost effect" is amplified for academic institutions that have little industry funding or that are relatively unknown in that particular field. Finally, the effect is not explained by heterogeneity among twin papers or by formal networks. Our finding that firms pay more attention to academic papers authored in hubs of industrial R&D does not mean that firms are irrational in doing so. After all, it is possible that the cost of "looking beyond the lampposts" is not worth the return. However, it is also possible that firms might be able to spot hidden opportunities by paying attention to discoveries emerging outside of hubs.

Our results complement the current literature on hubs (Saxenian 1994; Ellison, Glaeser, and Kerr 2010) by highlighting that their influence goes beyond their borders. We also depart from the emphasis on the stickiness of academic knowledge characteristic of prior work on science-based invention (Cockburn and Henderson 1998; Zucker, Darby, and Armstrong 2002) and highlight instead the breadth of the scientific literature and firms' challenge in allocating their scarce attention to it. Perhaps most importantly, our analysis indicates that elements external to a firm can shape what it does or does not pay attention to.

The benefits in building a richer understanding of external drivers of attention should not be understated. Firms' selective attention defines the menu of available opportunities that they consider. Innovation performance will suffer if they pay too much attention to unpromising opportunities and too little attention to promising ones. By underlining the effect of hubs as "lampposts" attracting firms' attention, we hope to highlight the fact that a firm's absorption of external knowledge depends not only on its internal capacity, but also on the salience of the external opportunities to which it has access.

References

Adams, James D. 2002. "Comparative Localization of Academic and Industrial Spillovers." *Journal of Economic Geography* 2 (3):253–278.

Agrawal, Ajay, Avi Goldfarb, and Florenta Teodoridis. 2016. "Understanding the Changing Structure of Scientific Inquiry." *American Economic Journal: Applied Economics* 8 (1):100–128. https://doi.org/10.1257/app.20140135.

Agrawal, Ajay, and Rebecca M. Henderson. 2002. "Putting Patents in Context: Exploring Knowledge Transfer from MIT." *Management Science* 48 (1).

Alcácer, Juan, and Michelle Gittelman. 2006. "Patent Citations as a Measure of Knowledge Flows: The Influence of Examiner Citations." *Review of Economics and Statistics* 88 (4):774–79. https://doi.org/i: 10.1162/rest.88.4.774</p>.

Alcácer, Juan, Michelle Gittelman, and Bhaven Sampat. 2009. "Applicant and Examiner Citations in U.S. Patents: An Overview and Analysis." *Research Policy* 38 (2):415–27. https://doi.org/16/j.respol.2008.12.001.

Azoulay, Pierre, Joshua S. Graff Zivin, and Bhaven N. Sampat. 2012. "The Diffusion of Scientific Knowledge Across Time and Space: Evidence from Professional Transitions for the Superstars of Medicine." In *The Rate of Direction of Inventive Activity*, 107–155. University of Chicago Press. http://www.nber.org/papers/w16683.

Belenzon, Sharon, and Mark Schankerman. 2013. "Spreading the Word: Geography, Policy, and Knowledge Spillovers." *Review of Economics and Statistics* 95 (3):884–903. https://doi.org/10.1162/REST_a_00334.

Bikard, Michaël. 2012. "Simultaneous Discoveries as a Research Tool: Method and Promise." *MIT Sloan Working Paper*.

Boudreau, Kevin J., Tom Brady, Ina Ganguli, Patrick Gaule, Eva Guinan, Anthony Hollenberg, and Karim R. Lakhani. 2017. “A Field Experiment on Search Costs and the Formation of Scientific Collaborations.” *The Review of Economics and Statistics*, April. https://doi.org/10.1162/REST_a_00676.

Branstetter, Lee, and Yoshiaki Ogura. 2005. “Is Academic Science Driving a Surge in Industrial Innovation? Evidence from Patent Citations.” National Bureau of Economic Research, Inc. RePEc. http://ideas.repec.org/p/nbr/nberwo/11561.html.

Callaert, Julie, Bart Van Looy, Arnold Verbeek, Koenraad Debackere, and Bart Thijs. 2006. “Traces of Prior Art: An Analysis of Non-Patent References Found in Patent Documents.” *Scientometrics* 69 (1):3–20. https://doi.org/10.1007/s11192-006-0135-8.

Carpenter, Mark P., and Francis Narin. 1983. “Validation Study: Patent Citations as Indicators of Science and Foreign Dependence.” *World Patent Information* 5 (3):180–85. https://doi.org/10.1016/0172-2190(83)90139-4.

Cockburn, Iain, and Rebecca Henderson. 1998. “Absorptive Capacity, Coauthoring Behavior, and the Organization of Research in Drug Discovery.” *The Journal of Industrial Economics* 46 (2):157–82.

Cohen, Wesley, and Daniel Levinthal. 1990. “Absorptive Capacity: A New Perspective on Learning and Innovation.” *Administrative Science Quarterly* 35 (1):128–52. https://doi.org/10.2307/2393553.

Cohen, Wesley, Richard Nelson, and John Walsh. 2002. “Links and Impacts: The Influence of Public Research on Industrial R&D.” *Management Science* 48 (1):1–23.

Cozzens, Susan E. 1989. *Social Control and Multiple Discovery in Science: The Opiate Receptor Case*. State University of New York Press.

Cyert, Richard M., and James G. March. 1963. *A Behavioral Theory of the Firm*. Prentice Hall.

Dahl, Michael S., and Christian Ø. R. Pedersen. 2004. “Knowledge Flows through Informal Contacts in Industrial Clusters: Myth or Reality?” *Research Policy* 33 (10):1673–86. https://doi.org/10.1016/j.respol.2004.10.004.

Drahl, Carmel. 2014. “Consecutive Journal Publications Illuminate Collaboration And Compromise In Chemistry.” *Chemical & Engineering News* 92 (40). http://cen.acs.org/articles/92/i40/Consecutive-Journal-Publications-Illuminate-Collaboration.html.

Ellison, Glenn, Edward L. Glaeser, and William R. Kerr. 2010. “What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns.” *American Economic Review* 100 (3):1195–1213. https://doi.org/10.1257/aer.100.3.1195.

Fleming, Lee, and Olav Sorenson. 2004. “Science as a Map in Technological Search.” *Strategic Management Journal* 25:909–28.

Furman, Jeffrey, and Megan J. MacGarvie. 2007. “Academic Science and the Birth of Industrial Research Laboratories in the U.S. Pharmaceutical Industry.” *Journal of Economic Behavior & Organization* 63 (4):756–76. https://doi.org/10.1016/j.jebo.2006.05.014.

Glaeser, Edward L., Hedi D. Kallal, José A. Scheinkman, and Andrei Shleifer. 1992. “Growth in Cities.” *Journal of Political Economy* 100 (6):1126–52. https://doi.org/10.1086/261856.

Halim, Nadia. 2000. “Bridging Apoptotic Signaling Gaps.” *The Scientist,* August 21, 2000. http://www.the-scientist.com/?articles.view/articleNo/12974/title/Bridging-Apoptotic-Signaling-Gaps/.

Hansen, Morten T., and Martine R. Haas. 2001. “Competing for Attention in Knowledge Markets: Electronic Document Dissemination in a Management Consulting Company.” *Administrative Science Quarterly* 46 (1):1–28. https://doi.org/10.2307/2667123.

Henderson, Rebecca, Adam Jaffe, and Manuel Trajtenberg. 2005. “Patent Citations and the Geography of Knowledge Spillovers: A Reassessment: Comment.” *The American Economic Review* 95 (1):461–64.

Jacobs, Jane. 1970. *The Economy of Cities*. New York: Vintage.

Jaffe, Adam. 1989. “Real Effects of Academic Research.” *The American Economic Review* 79 (5):957–70.

Jaffe, Adam, Manuel Trajtenberg, and Rebecca Henderson. 1993. “Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations.” *The Quarterly Journal of Economics* 108 (3):577–98. https://doi.org/10.2307/2118401.

Jinha, Arif E. 2010. “Article 50 Million: An Estimate of the Number of Scholarly Articles in Existence.” *Learned Publishing* 23 (3):258–63. https://doi.org/10.1087/20100308.

Jones, Benjamin. 2009. “The Burden of Knowledge and the ‘Death of the Renaissance Man’: Is Innovation Getting Harder?” *Review of Economic Studies* 76 (1):283–317.

Lampe, Ryan. 2012. “Strategic Citation.” *Review of Economics and Statistics* 94 (1):320–33. https://doi.org/10.1162/REST_a_00159.

Lemley, Mark A. 2007. “Should Patent Infringement Require Proof of Copying?” *Michigan Law Review* 105 (7):1525–36.

Li, Guan-Cheng, Ronald Lai, Alexander D’Amour, David M. Doolin, Ye Sun, Vetle I. Torvik, Amy Z. Yu, and Lee Fleming. 2014. “Disambiguation and Co-Authorship Networks of the U.S. Patent Inventor Database (1975–2010).” *Research Policy* 43 (6):941–55. https://doi.org/10.1016/j.respol.2014.01.012.

Li, Qiang, Patrick G. Maggitti, Ken G. Smith, Paul E. Tesluk, and Riitta Katila. 2013. “Top Management Attention to Innovation: The Role of Search Selection and Intensity in New Product

Introductions." *Academy of Management Journal* 56 (3):893–916. https://doi.org/10.5465/amj.2010.0844.

March, James G. 1991. "Exploration and Exploitation in Organizational Learning." *Organization Science* 2 (1):71–87. https://doi.org/10.2307/2634940.

March, James G., and Herbert Alexander Simon. 1958. "Organizations."

Marshall, A. 1895. *Principles of Economics*. Macmillan.

Merton, Robert K. 1961. "Singletons and Multiples in Scientific Discovery: A Chapter in the Sociology of Science." *Proceedings of the American Philosophical Society* 105 (5):470–86.

Mill, John Stuart. 1848. *Principles of Political Economy with Some of Their Applications to Social Philosophy*. JW Parker.

Monteiro, L. Felipe. 2015. "Selective Attention and the Initiation of the Global Knowledge-Sourcing Process in Multinational Corporations." *Journal of International Business Studies* 46 (5):505–527.

Murray, Fiona. 2002. "Innovation as Co-Evolution of Scientific and Technological Networks: Exploring Tissue Engineering." *Research Policy* 31 (8–9):1389–1403.

Narin, Francis, Kimberly S. Hamilton, and Dominic Olivastro. 1997. "The Increasing Linkage between U.S. Technology and Public Science." *Research Policy* 26 (3):317–30. https://doi.org/16/S0048-7333(97)00013-9.

Niehans, Jurg. 1995. "Multiple Discoveries in Economic Theory." *European Journal of the History of Economic Thought* 2 (1):1. https://doi.org/Article.

Ocasio, William. 1997. "Towards an Attention-Based View of the Firm." *Strategic Management Journal* 18:187–206.

———. 2010. "Attention to Attention." *Organization Science* 22 (5):1286–96. https://doi.org/10.1287/orsc.1100.0602.

Padgett, John F., and Paul D. McLean. 2006. "Organizational Invention and Elite Transformation: The Birth of Partnership Systems in Renaissance Florence." *American Journal of Sociology* 111 (5):1463–1568. https://doi.org/10.1086/498470.

Piezunka, Henning, and Linus Dahlander. 2015. "Distant Search, Narrow Attention: How Crowding Alters Organizations' Filtering of Suggestions in Crowdsourcing." *Academy of Management Journal* 58 (3):856–80. https://doi.org/10.5465/amj.2012.0458.

Roach, Michael, and Wesley M. Cohen. 2013. "Lens or Prism? Patent Citations as a Measure of Knowledge Flows from Public Research." *Management Science* 59 (2):504–25. https://doi.org/10.1287/mnsc.1120.1644.

Saxenian, AnnaLee. 1994. *Regional Advantage: Culture and Competition in Silicon Valley and Route 128*. Harvard University Press.

Simon, Herbert. 1947. *Administrative Behavior*. Free Press. http://www.amazon.ca/exec/obidos/redirect?tag=citeulike09-20&path=ASIN/0684835827.

Singh, Jasjit. 2005. "Collaborative Networks as Determinants of Knowledge Diffusion Patterns." *Management Science* 51 (5):756–70. https://doi.org/10.1287/mnsc.1040.0349.

Singh, Jasjit, and Matt Marx. 2013. "Geographic Constraints on Knowledge Spillovers: Political Borders vs. Spatial Proximity." *Management Science* 59 (9):2056–78. https://doi.org/10.1287/mnsc.1120.1700.

Smith, Adam. 1776. *An Inquiry into the Nature and Causes of the Wealth of Nations*.

Sohn, Eunhee. 2014. "The Endogeneity of Academic Science to Local Industrial R&D." *Academy of Management Proceedings* 2014 (1):11413. https://doi.org/10.5465/AMBPP.2014.286.

Stigler, George J. 1980. "Merton on Multiples, Denied and Affirmed." *Transactions of the New York Academy of Sciences* 39 (1 Series II):143–46. https://doi.org/10.1111/j.2164-0947.1980.tb02774.x.

Thompson, Peter, and Melanie Fox-Kean. 2005. "Patent Citations and the Geography of Knowledge Spillovers: A Reassessment." *The American Economic Review* 95 (1):450–60.

Vermont, Samson. 2006. "Independent Invention as a Defense to Patent Infringement." *Michigan Law Review* 105 (3):475–504.

Weintraub, E. Roy. 2011. "Retrospectives: Lionel W. McKenzie and the Proof of the Existence of a Competitive Equilibrium." *Journal of Economic Perspectives* 25 (2):199–215. https://doi.org/10.1257/jep.25.2.199.

Zucker, Lynne G., Michael R. Darby, and Jeff S. Armstrong. 2002. "Commercializing Knowledge: University Science, Knowledge Capture, and Firm Performance in Biotechnology." *Management Science* 48 (1):138–53.

Zucker, Lynne G., Michael R. Darby, and Marilynn B. Brewer. 1998. "Intellectual Human Capital and the Birth of U.S. Biotechnology Enterprises." *The American Economic Review* 88 (1):290–306.

Table I: Summary statistics for twin academic papers reporting simultaneous discoveries. N=1,166.

		Mean	Std. Dev.	Min.	Max.
Academic "twin" papers from simultaneous discoveries where no paper was referenced by any corporate patent (N=786)	Number of patents referencing twin paper	0.000	0.000	0	0
	Journal impact factor	21.889	12.047	0	51.410
	Paper located in U.S.	0.604	0.489	0	1
	Paper was patented	0.103	0.304	0	1
	Corresponding author stock of patents	1.042	3.541	0	55
	Corresponding author stock of papers	74.005	87.304	0	754
	Institution's 5-year stock of patents	181.315	321.256	0	2308
	Institution's prestige in the paper's field	17.292	30.957	0	176
Academic "twin" papers from simultaneous discoveries where twins were always referenced jointly (N=18)	Number of patents referencing twin paper	1.389	0.502	1	2
	Paper author(s) in a hub of relevant R&D	0.222	0.428	0	1
	Journal impact factor	20.468	11.019	2.19	32.906
	Paper located in U.S.	0.500	0.514	0	1
	Paper was patented	0.222	0.428	0	1
	Corresponding author stock of patents	1.056	2.071	0	6
	Corresponding author stock of papers	50.222	54.051	9	192
	Institution's 5-year stock of patents	144.041	184.729	0	584
	Institution's prestige in the paper's field	23.389	44.942	0	176
Academic "twin" papers from simultaneous discoveries where twins were sometimes cited jointly (N=158)	Number of patents referencing twin paper	7.652	9.742	0	56
	Paper author(s) in a hub of relevant R&D	0.184	0.388	0	1
	Journal impact factor	25.724	10.594	4.18	51.410
	Paper located in U.S.	0.671	0.471	0	1
	Paper was patented	0.316	0.467	0	1
	Corresponding author stock of patents	1.816	6.849	0	75
	Corresponding author stock of papers	82.532	110.840	0	640
	Institution's 5-year stock of patents	111.670	179.352	0	1394
	Institution's prestige in the paper's field	21.281	35.285	0	176
Academic "twin" papers from simultaneous discoveries where twins were cited but never jointly (N=185)	Number of patents referencing twin paper	1.962	4.168	0	37
	Paper author(s) in a hub of relevant R&D	0.108	0.311	0	1
	Journal impact factor	22.412	11.466	0	51.410
	Paper located in U.S.	0.643	0.480	0	1
	Paper was patented	0.141	0.348	0	1
	Corresponding author stock of patents	1.341	3.442	0	25
	Corresponding author stock of papers	72.735	83.456	0	560
	Institution's 5-year stock of patents	170.480	247.137	0	1415
	Institution's prestige in the paper's field	20.102	33.885	0	176

Table II: Location of academic publications belonging to sets of "paper twins" where one but not all papers are referenced

Panel A

Institutions with four or more "twin" academic papers

	Freq.	Percent	Cum.
Harvard University	15	4.79	4.79
UT Southwestern Medical Ctr	11	3.51	8.31
UC San Francisco	9	2.88	11.18
Columbia University	8	2.56	13.74
Johns Hopkins University	8	2.56	16.29
MIT	8	2.56	18.85
Salk Institute	7	2.24	21.09
Rockefeller University	7	2.24	23.32
University of Toronto	6	1.92	25.24
Yale University	6	1.92	27.16
UC San Diego	5	1.6	28.75
Oxford University	5	1.6	30.35
European Molecular Biology Lab	4	1.28	31.63
London Research Institute	4	1.28	32.91
Massachusetts Gen. Hospital	4	1.28	34.19
RIKEN	4	1.28	35.46
Cambridge University	4	1.28	36.74
Duke University	4	1.28	38.02
University of North Carolina	4	1.28	39.3
New York University	4	1.28	40.58
Stanford University	4	1.28	41.85
University of Washington	4	1.28	43.13

Panel B

Cities with four or more "twin" academic papers

	Freq.	Percent	Cum.
Boston, MA	26	8.31	8.31
New York, NY	23	7.35	15.65
San Diego, CA	13	4.15	19.81
Bethesda, MD	10	3.19	23
San Francisco, CA	9	2.88	25.88
Baltimore, MD	8	2.56	28.43
Cambridge, MA	8	2.56	30.99
Dallas, TX	8	2.56	33.55
London, UK	7	2.24	35.78
New Haven, CT	7	2.24	38.02
Toronto, Canada	7	2.24	40.26
Cambridge, UK	6	1.92	42.17
Heidelberg, Germany	5	1.6	43.77
Houston, TX	5	1.6	45.37
Oxford, UK	5	1.6	46.96
Philadelphia, PA	5	1.6	48.56
Seattle, WA	5	1.6	50.16
Chapel Hill, NC	4	1.28	51.44
Chicago, IL	4	1.28	52.72
Durham, NC	4	1.28	53.99
Los Angeles, CA	4	1.28	55.27
Palo Alto, CA	4	1.28	56.55

Notes: "Twin" papers report the same simultaneous academic discovery. The subset of 316 twin papers here are limited to simultaneous discoveries where one but not all twins was referenced by a firm-assigned patent.

Table III: Summary statistics for academic-paper/industry-patent dyads

	Obs	Mean	Std. Dev	Min	Max
Twin paper referenced by focal patent	1671	0.478	0.500	0	1
Paper author(s) in a hub	1671	0.244	0.430	0	1
Ln # of patents within 50 miles of corresponding author	1671	0.610	1.225	0	5.628
Ln # conferences in this field held within 50 miles	1671	3.703	1.661	0	6.033
Author(s) near hub * >median # of conferences	1671	0.235	0.424	0	1
Author(s) near hub * <median # of conferences	1671	0.010	0.097	0	1
Paper published before its twin(s)	1671	0.202	0.402	0	1
Paper more detailed than its twin(s)	552	0.408	0.492	0	1
Paper has more claims than its twin(s)	552	0.219	0.414	0	1
Paper written more clearly than its twin(s)	552	0.353	0.478	0	1
Paper more clinical than its twin(s)	552	0.069	0.253	0	1
Coauthor of a paper author is an inventor on patent	1671	0.155	0.362	0	1
Ln journal impact factor	1671	3.206	0.577	0	3.959
Paper located in U.S.	1671	0.668	0.471	0	1
Paper was patented	1671	0.245	0.430	0	1
Ln corresponding author stock of patents	1671	0.453	0.749	0	4.331
Ln corresponding author stock of papers	1671	3.567	1.343	0	6.463
Institution's 5-year stock of patents	1671	110.619	150.849	0	1415
Institution's prestige in the paper's field	1671	21.742	35.540	0	176
Ln population of corresponding author's city	1671	13.108	1.637	8.531	16.8
Publication lag, paper vs. patent	1671	5.135	3.215	0	17
Ln distance between paper and patent	1671	7.129	2.020	0	9.263
Paper and patent in same state	1671	0.101	0.301	0	1
Paper and patent in same country	1671	0.517	0.500	0	1

Notes: Observations are constructed for all combinations of twin academic papers and patents where one but not all twin academic papers reporting a simultaneous discovery is referenced by a firm-assigned patent.

Table IV: The impact of the location of academic institutions on paper citation in patents

	Dependent variable indicates that the "twin" paper was referenced by a patents of the following type				
	industry				**academic**
	(1)	(2)	(3)	(4)	(5)
Paper author(s) in a hub		0.717*	1.024*		0.496
		(0.316)	(0.514)		(0.330)
Ln # patents within 50 miles				0.929***	
of corresponding author				(0.183)	
Ln distance between paper and patent	-0.244***	-0.171**	-0.108	-0.157**	0.0197
	(0.0626)	(0.0559)	(0.0916)	(0.0563)	(0.0751)
Observations	1,671	1,671	1,671	1,671	1,088
# twin articles	316	316	316	316	381
Pseudo-R2	0.065	0.162	0.490	0.173	0.0554
Log-likelihood	-547.4	-490.7	-298.8	-484.3	-359.6
paper, author, institution controls	no	yes	yes	yes	yes
City fixed effects for principal investigat	no	no	yes	no	no
Simultaneous-discovery/patent FE	yes	yes	yes	yes	yes

Notes: Observations are academic-paper/industry-patent dyads. All models are estimated using conditional logit and include simultaneous-discovery/patent fixed effects. Controls include those for the paper (U.S.-based, journal impact factor, paper was patented), corresponding author (stock of patents and papers, population of city), and institution (stock of patents and papers) characteristics as well as characteristics of the paper-patent dyad (publication lag). Standard errors are clustered at the level of the simultaneous discovery; *** p<0.001; ** p<0.01; * p<.0.05; + p<0.1.

Table V: Difference of means tests for within-twin heterogeneity

	Papers in hubs (n=41)	Papers not in hubs (N=45)	*p* <
Paper is more detailed	0.390	0.311	0.447
Paper has broader claims	0.220	0.067	0.042
Paper is more clearly written	0.414	0.266	0.151
Paper is more clinically oriented	0.073	0.067	0.907

Notes: Results of expert-panel evaluation of 86 "twin" papers by ten postdoctoral researchers in the life sciences and applied physics. Postdocs identified the four criteria in the first column and classified twin papers accordingly.

Table VI: Mechanisms

	(1)	(2)	(3)	(4)
Paper author(s) in a hub	1.575***	0.716*	0.641*	
	(0.467)	(0.312)	(0.307)	
Paper published before its twin(s)	-0.285			
	(0.609)			
Paper more detailed than its twin(s)	0.382			
	(0.604)			
Paper has more claims than its twin(s)	-0.849			
	(0.608)			
Paper written more clearly than its twin(s)	0.523			
	(0.416)			
Paper more clinical than its twin(s)	1.346+			
	(0.688)			
Coauthor of a paper author is an inventor on patent		0.0363		
		(0.391)		
Ln # conferences in this field held within 50 miles of corresponding author			0.150+	
			(0.0851)	
Author(s) in hub * >median # of conferences				0.733*
				(0.330)
Author(s) in hub * <median # of conferences				0.460
				(0.908)
Ln distance between paper and patent	-0.208	-0.171**	-0.137**	-0.170**
	(0.163)	(0.0566)	(0.0509)	(0.0565)
Observations	552	1,671	1,671	1,671
# twin articles	78	316	316	316
Pseudo-R2	0.260	0.162	0.170	0.162
Log-likelihood	-143.4	-490.7	-485.9	-490.7
paper, author, institution controls	yes	yes	yes	yes
Simultaneous-discovery/patent FE	yes	yes	yes	yes

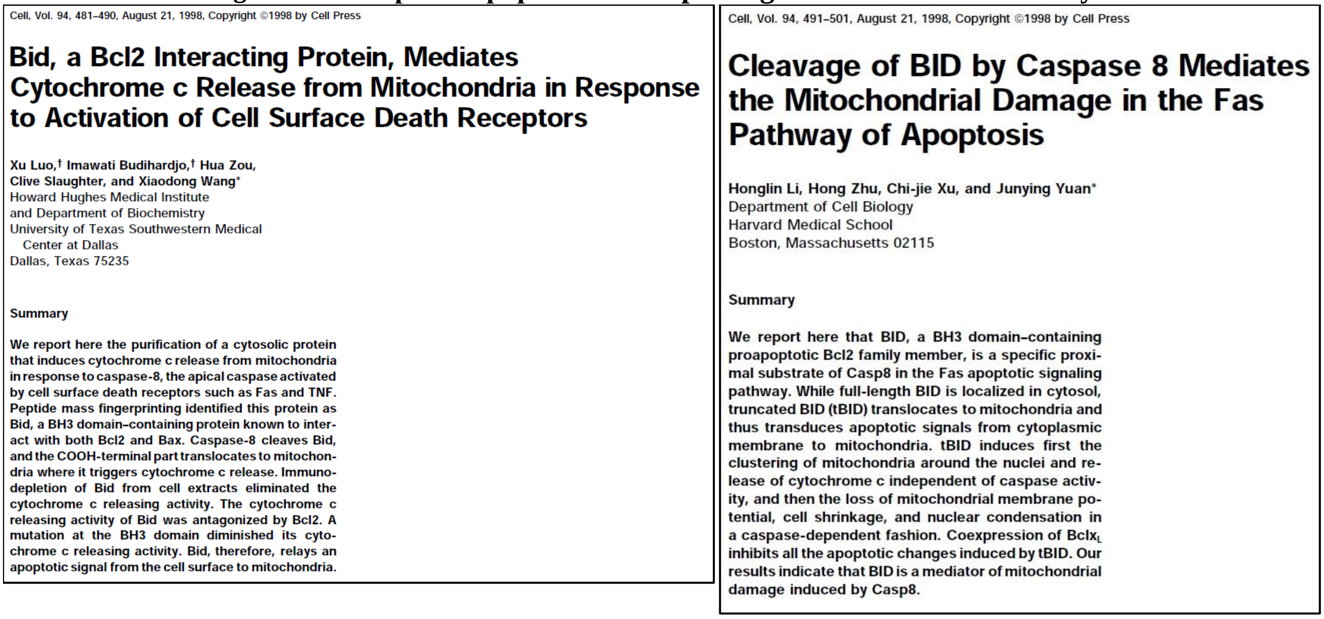
Notes: All models include controls for characteristics of the paper (U.S.-based, journal impact factor, discovery was patented), author (stock of patents and papers, city population), institution (stock of patents and papers), and paper-patent dyad (publication lag; spatial distance). Column (1) contains fewer observations as its analysis is limited to the twin papers evaluated for heterogeneity by the expert panel. Standard errors are clustered throughout at the level of the simultaneous discovery; *** p<0.001; ** p<0.01; * p<.0.05; + p<0.1.

Table VII: Institutional Heterogeneity

Panel A Industry investment in the focal institution	(1)	**Panel B** Institutional prestige	(2)	**Panel C** Distance between focal paper and patent	(3)
In a hub & no industry funding	1.645*	In a hub & below-median prestige	2.250**	Near a hub & within 20 miles of patent	0.152
	(0.749)		(0.704)		(0.761)
In a hub & little industry funding	3.810*	In a hub & above-media prestige	0.286	Near a hub & within 20-2000 miles of patent	1.582***
	(1.758)		(0.325)		(0.437)
In a hub & more industry funding	0.835+			Near a hub & 2000+ miles away from patent	0.285
	(0.440)				(0.427)
In a hub & most industry funding	0.485				
	(0.657)				
Observations	898	Observations	1,671	Observations	1,671
# twin articles	166	# twin articles	316	# twin articles	316
Pseudo-R2	0.264	Pseudo-R2	0.156	Pseudo-R2	0.173
Log-likelihood	-230.6	Log-likelihood	-494	Log-likelihood	-474.3
Simultaneous-discovery/patent FE	yes	Simultaneous-discovery/patent FE	yes	Simultaneous-discovery/patent FE	yes

Notes: Observations are academic-paper/industry-patent dyads. The dependent variable indicates whether the patent in the dyad references the paper. All models are estimated using simultaneous-discovery/patent fixed effects. All models include controls for the paper (U.S.-based, journal impact factor, discovery was patented), corresponding author (stock of patents and papers, population of city), and institution (stock of patents and papers) characteristics as well as characteristics of the paper-patent dyad (publication lag; spatial distance). Base variables for interactions are not shown. The omitted category for the interactions consists of papers that were located in hubs of commercial R&D in the same scientific field as the discovery. Panel A uses data from North America only due to the scope of data available from the Association for University Technology Managers. Standard errors are clustered at the level of the simultaneous discovery; *** p<0.01; ** p<0.1; * p<.0.05; + p<0.10.

Figure 1: Example of "paper twins" reporting a simultaneous discovery



Cell, Vol. 94, 481–490, August 21, 1998, Copyright ©1998 by Cell Press

Bid, a Bcl2 Interacting Protein, Mediates Cytochrome c Release from Mitochondria in Response to Activation of Cell Surface Death Receptors

Xu Luo,† Imawati Budihardjo,† Hua Zou, Clive Slaughter, and Xiaodong Wang*
Howard Hughes Medical Institute
and Department of Biochemistry
University of Texas Southwestern Medical
Center at Dallas
Dallas, Texas 75235

Summary

We report here the purification of a cytosolic protein that induces cytochrome c release from mitochondria in response to caspase-8, the apical caspase activated by cell surface death receptors such as Fas and TNF. Peptide mass fingerprinting identified this protein as Bid, a BH3 domain–containing protein known to interact with both Bcl2 and Bax. Caspase-8 cleaves Bid, and the COOH-terminal part translocates to mitochondria where it triggers cytochrome c release. Immunodepletion of Bid from cell extracts eliminated the cytochrome c releasing activity. The cytochrome c releasing activity of Bid was antagonized by Bcl2. A mutation at the BH3 domain diminished its cytochrome c releasing activity. Bid, therefore, relays an apoptotic signal from the cell surface to mitochondria.

Cell, Vol. 94, 491–501, August 21, 1998, Copyright ©1998 by Cell Press

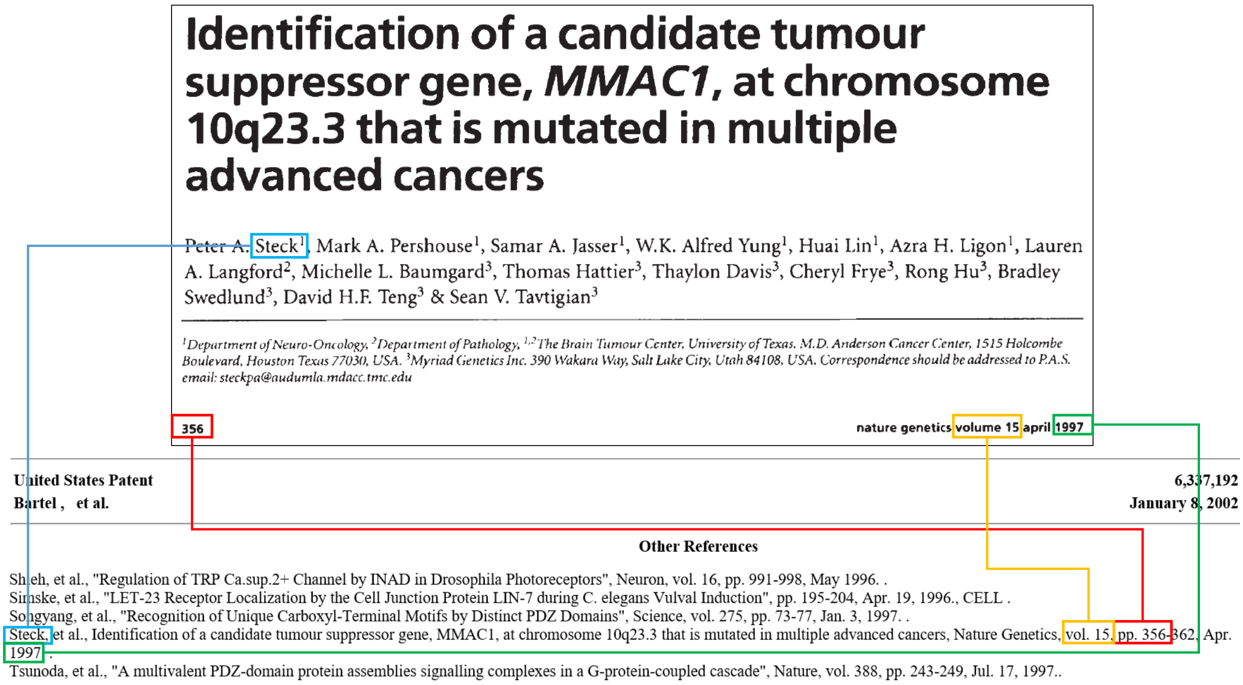
Cleavage of BID by Caspase 8 Mediates the Mitochondrial Damage in the Fas Pathway of Apoptosis

Honglin Li, Hong Zhu, Chi-jie Xu, and Junying Yuan*
Department of Cell Biology
Harvard Medical School
Boston, Massachusetts 02115

Summary

We report here that BID, a BH3 domain–containing proapoptotic Bcl2 family member, is a specific proximal substrate of Casp8 in the Fas apoptotic signaling pathway. While full-length BID is localized in cytosol, truncated BID (tBID) translocates to mitochondria and thus transduces apoptotic signals from cytoplasmic membrane to mitochondria. tBID induces first the clustering of mitochondria around the nuclei and release of cytochrome c independent of caspase activity, and then the loss of mitochondrial membrane potential, cell shrinkage, and nuclear condensation in a caspase-dependent fashion. Coexpression of $Bclx_L$ inhibits all the apoptotic changes induced by tBID. Our results indicate that BID is a mediator of mitochondrial damage induced by Casp8.

Figure II: Patent-to-paper References



Identification of a candidate tumour suppressor gene, *MMAC1*, at chromosome 10q23.3 that is mutated in multiple advanced cancers

Peter A. Steck[1], Mark A. Pershouse[1], Samar A. Jasser[1], W.K. Alfred Yung[1], Huai Lin[1], Azra H. Ligon[1], Lauren A. Langford[2], Michelle L. Baumgard[3], Thomas Hattier[3], Thaylon Davis[3], Cheryl Frye[3], Rong Hu[3], Bradley Swedlund[3], David H.F. Teng[3] & Sean V. Tavtigian[3]

[1]Department of Neuro-Oncology, [2]Department of Pathology, [1,2]The Brain Tumour Center, University of Texas, M.D. Anderson Cancer Center, 1515 Holcombe Boulevard, Houston Texas 77030, USA. [3]Myriad Genetics Inc. 390 Wakara Way, Salt Lake City, Utah 84108, USA. Correspondence should be addressed to P.A.S. email: steckpa@audumla.mdacc.tmc.edu

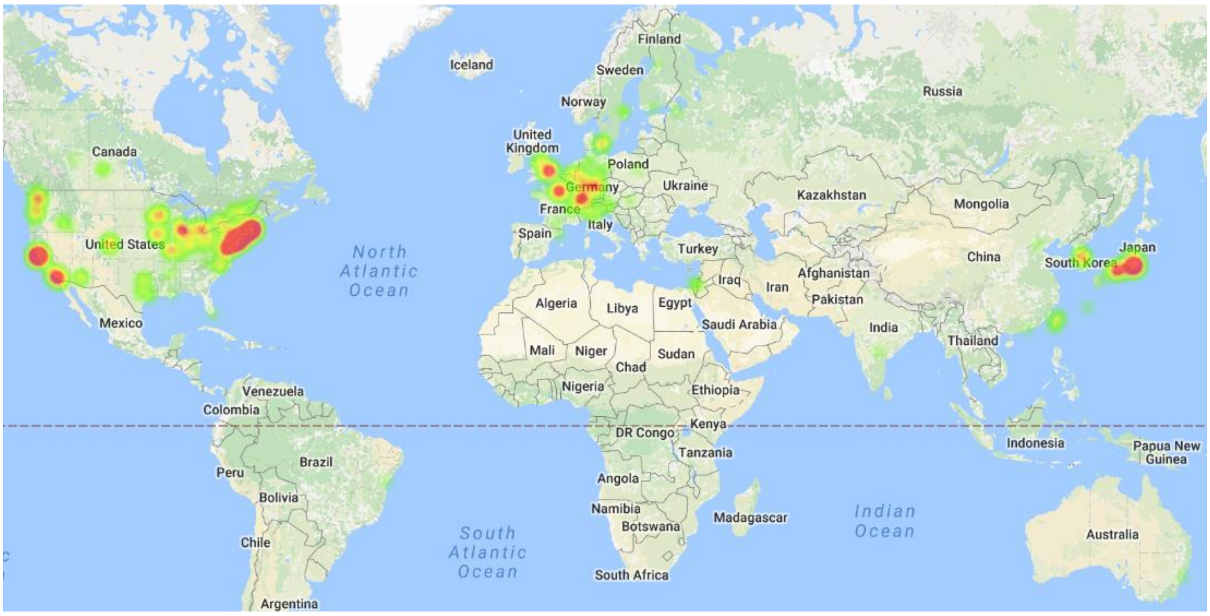
356 nature genetics volume 15 april 1997

United States Patent 6,337,192
Bartel , et al. January 8, 2002

Other References

Shieh, et al., "Regulation of TRP Ca.sup.2+ Channel by INAD in Drosophila Photoreceptors", Neuron, vol. 16, pp. 991-998, May 1996. .
Simske, et al., "LET-23 Receptor Localization by the Cell Junction Protein LIN-7 during C. elegans Vulval Induction", pp. 195-204, Apr. 19, 1996., CELL .
Songyang, et al., "Recognition of Unique Carboxyl-Terminal Motifs by Distinct PDZ Domains", Science, vol. 275, pp. 73-77, Jan. 3, 1997. .
Steck, et al., Identification of a candidate tumour suppressor gene, MMAC1, at chromosome 10q23.3 that is mutated in multiple advanced cancers, Nature Genetics, vol. 15, pp. 356-362, Apr. 1997 .
Tsunoda, et al., "A multivalent PDZ-domain protein assemblies signalling complexes in a G-protein-coupled cascade", Nature, vol. 388, pp. 243-249, Jul. 17, 1997..

Figure III: Hub heatmaps
Panel A: worldwide



Panel B: U.S. only

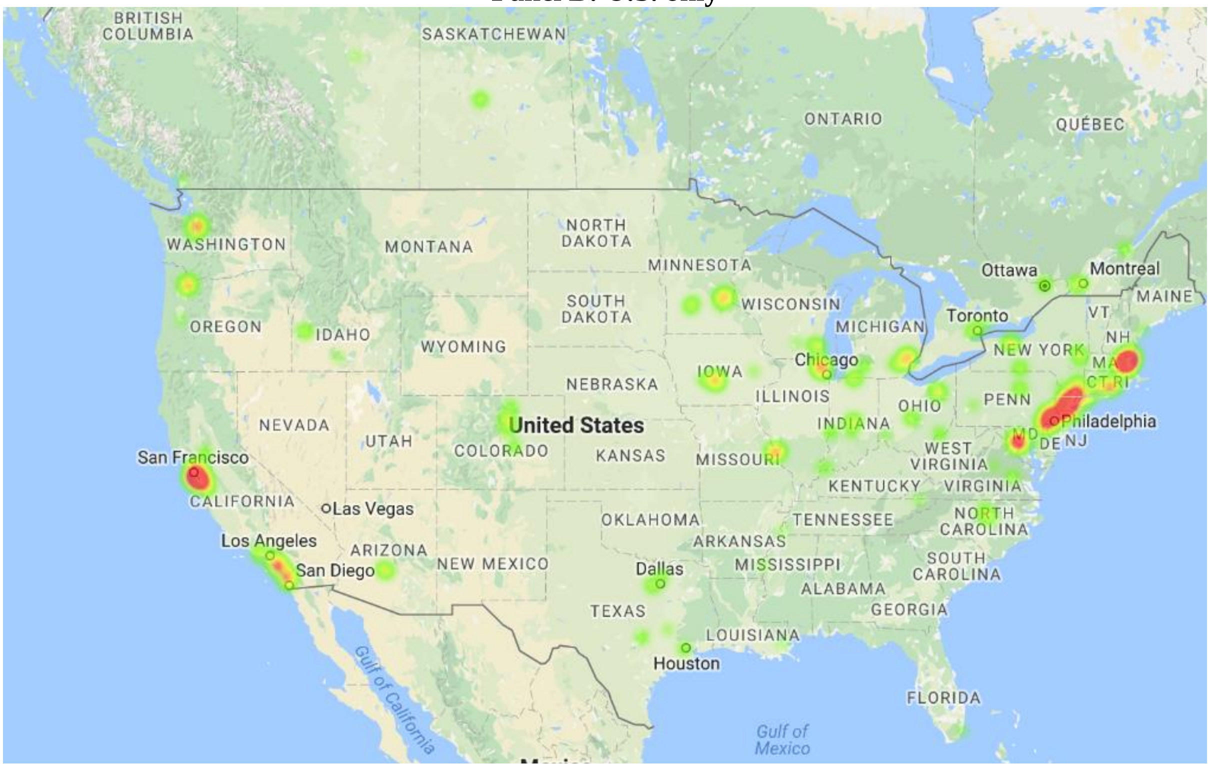
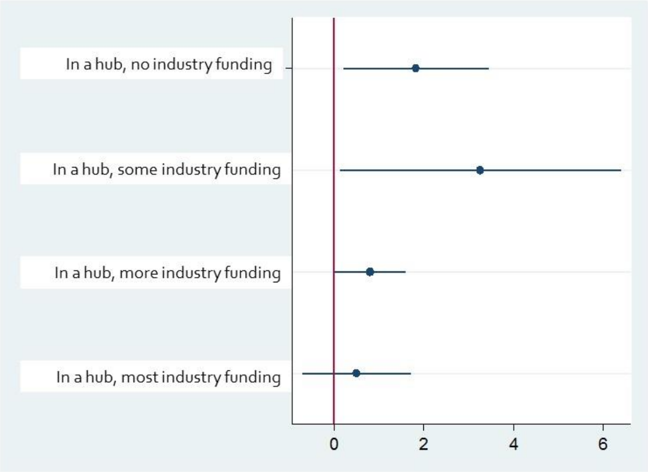
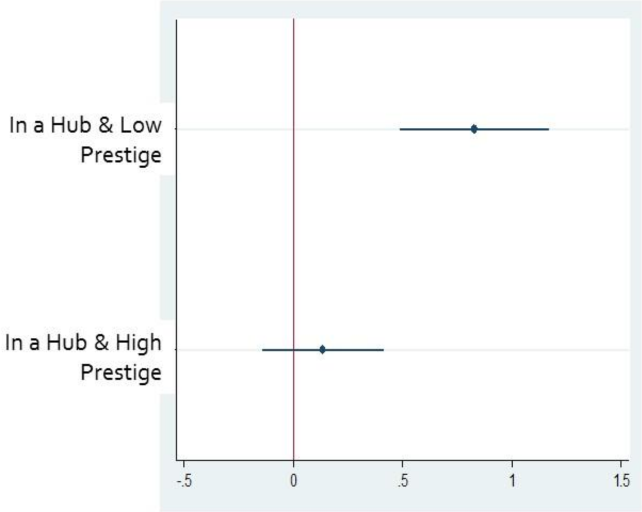


Figure IV: Institutional Heterogeneity

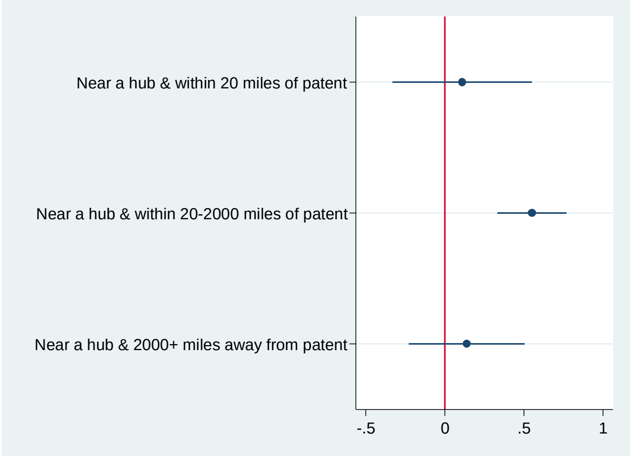
Panel A: Hubs and Funding of R&D at the paper's institution by industry



Panel B: Hubs and Organizational prestige



Panel C: Hubs and Distance between paper and patent



Notes: Coefficients are plotted from linear probability models.