The Location of Academic Institutions and Knowledge Flow to Industry: Evidence from Simultaneous Discoveries

Michaël Bikard
London Business School
mbikard@london.edu

Matt Marx
MIT Sloan School of Management
mmarx@mit.edu

Abstract: Scientific discoveries in academia can spur innovation and economic growth, but only if they flow to industry. This paper documents a source of friction in the flow of academic science to firms: corporate inventors tend to overlook academic discoveries that emerge outside concentrations or "hubs" of commercial R&D in the same particular field. Testing the impact of location on knowledge flow is difficult because institutions at different locations produce different kinds of research. We address this problem by analyzing simultaneous discoveries where multiple researchers publish "twin" papers which report the same finding. Even after accounting for the localization of knowledge flows, we find that a twin paper conducted outside of a hub of relevant R&D is approximately 10% less likely to be referenced as prior art by firm-assigned patents. This effect is moderated by collocation with the focal patent, the institution's academic prestige, and by formal connections with industry. Taken together, our results suggest that the geographic location of academic institutions affects the chances that their discoveries become orphaned, with sobering implications for the science of science policy yet strategic opportunities for firms.

JEL codes: O00, O32

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

Academic research is an essential engine of innovation and growth (Romer 1990; Grossman and Helpman 1993; Aghion, Dewatripont, and Stein 2008), with governments across the world investing billions annually in the hope that economic benefits will follow. Although academic research can increase an industry's R&D efficiency (Nelson 1959; Nelson 1982; Cohen, Nelson, and Walsh 2002; Mokyr 2002), such benefits accrue only if this knowledge flows to firms. Such concerns are not just theoretical: NIH director Francis Collins declared that he was "frustrated to see how many of the [academic] discoveries that do look as though they have therapeutic implications are waiting for the pharmaceutical industry to follow through with them" (Harris 2011). This observation reflects the need for greater understanding of the circumstances under which valuable academic knowledge fails to be utilized in the private sector.

The flow of scientific knowledge to industry may be affected by the location of the academic research institution where it originates. Since much knowledge flow is localized, location might affect knowledge flow because of the geographic distance that separates academic research institutions from firms (Jaffe, Trajtenberg, and Henderson 1993; Zucker, Darby, and Brewer 1998; Adams 2002; Thompson 2006; Furman and MacGarvie 2007; Azoulay, Graff Zivin, and Sampat 2012; Belenzon and Schankerman 2013). However, the impact of location on knowledge flow may not be limited to geographic distance between the source and recipient of the knowledge. We suggest that knowledge is likely to circulate more widely when it emerges in locations that are central to the community of commercial inventors. Conversely, discoveries from less central locations are less likely to be utilized.

Consider a San Diego-based biotech firm building on scientific knowledge discovered simultaneously by academic researchers in both Dallas, TX and Boston, MA. One might expect that the biotech firm is more likely to utilize the Dallas discovery since it is closer. However, we predict that the firm will exploit the other discovery because Boston is more a "hub" of biotech R&D. Academic research conducted in commercial R&D hubs in that same field is likely to receive more exposure than the same research conducted outside such a hub. Not only are corporate inventors likely to monitor new findings coming out of those hubs, but they are also likely to be familiar with the individuals and firms that work there. Hence, the location of an academic institution inside a hub can offset distance as a constraint on diffusion, and valuable knowledge developed outside of such hubs is at a higher risk of being ignored.

Indeed, Appendix I suggests that academic discoveries published even in top scientific journals are less likely to flow to industry (as measured by references from patents assigned to firms, see Appendix III) when the paper is published at an institution that is outside a hub of commercial R&D in that specific scientific field (see Appendix IV). Studying this phenomenon in a large sample of academic publications referenced by corporate patents affords a measure of external validity, but this approach is vulnerable to the criticism that scientific discoveries conducted inside hubs of commercial R&D may differ fundamentally from those emerging from outside such hubs. Industry inventors might ignore a

scientific discovery not because of its location but because it is less "applied" or simply of lower quality. There may in fact be a causal loop between industry demand for specific research and academic output; the direction of academic science is in part endogenous to local firms' research priorities (Sohn 2014). Therefore, although we are interested in the marginal impact of location of the academic institution on knowledge flow to industry, a selection effect could confound inference even if academic knowledge more relevant to industry is seen to emerge in institutions that happen to be located in R&D hubs.

To address these issues, we exploit the occurrence of simultaneous scientific discoveries (Merton 1961)[1]. When two or more scholars publish their findings at about the same time, they create "paper twins." By embodying a single piece of knowledge that emerged in multiple locations, paper twins control for the nature of the underlying science. We measure the flow of academic discoveries to industry by observing references from firm-assigned patents to paper twins where one twin is within a hub of relevant R&D and the other is not. We thus build on a large literature using references in patents and publications as a measure of knowledge flow (Jaffe, Trajtenberg, and Henderson 1993; Griliches 1998; Furman and Stern 2011; Galasso and Schankerman 2014).

Our empirical strategy has three key advantages. First, the use of patent references as a measure of knowledge flow is usually complicated by the possible existence of false positives, as citations are often added ex-post for legal or strategic reasons (Alcácer, Gittelman, and Sampat 2009; Lampe 2012). In the case of paper twins, however, our identification strategy is facilitated by USPTO Rule 56, which states that an inventor is *not required to reference multiple sources disclosing the same prior art*. Thus if a simultaneous discovery were relevant prior art, but the patent referenced only one of the "twin" papers reporting that discovery, failing to reference another twin paper would not affect the validity of the patent. This rule opens the possibility for a patent to reference one member of a twin set but not the other, allowing our study to focus on the question of why that twin was the one referenced.

Second, since academic discoveries tend to be published and not patented, a focus on references to academic patents could introduce a bias by excluding the majority of academic discoveries. We circumvent this difficulty by using patent applicants' references not to other patents but to scientific articles (e.g., Belenzon and Schankerman 2013; Azoulay, Graff Zivin, and Sampat 2012). While references from patents to scientific publications by no means capture every flow of academic knowledge to commercial R&D, Roach and Cohen (2013) report that they may be the most reliable indicator.

Third, as in any case-control analysis, inference depends on the similarity between treatment and

[1] We focus on natural sciences, but simultaneous discoveries also occur elsewhere such as the existence of a competitive equilibrium in a market economy by McKenzie and by Arrow-Debreu in 1954 (Weintraub 2011); other cases have also been reported (Stigler 1980; Niehans 1995).

controls (Thompson and Fox-Kean 2005). Drawing inferences regarding the influence of location on the flow of knowledge from academia to industry is challenging because discoveries in multiple locations may differ in several respects. As noted above, an academic paper might appear to receive less exposure to industry because it is in a remote location, but perhaps this is due to the discovery being more basic, or theoretical, and hence of less interest to firms. To the extent that our "twin" papers truly represent the same discovery, they provide an opportunity to control for the quality of the underlying discovery.

By examining patent references to twin papers, we replicate the established finding that the flow of academic knowledge to industry attenuates with spatial separation between the academic scientist and the firm (Jaffe, Trajtenberg, and Henderson 1993; Belenzon and Schankerman 2013; Mowery and Ziedonis 2015). We also show that controlling for localization, academic discoveries made outside of relevant commercial R&D hubs are less likely to flow to inventors working in firms. This result is robust to a number of specifications and is only visible among corporate (not academic) inventors. However, this effect is attenuated in the following circumstances: 1) for institutions with formal connections to industry, 2) for papers at institutions with a higher academic reputation, and 3) when the focal paper and the potentially-referencing patent are themselves collocated. Thus it appears that the negative impact of being outside of an R&D hub is mitigated when commercial inventors are exposed to a paper by other means.

Our findings suggest that public investment in academic research outside of relevant hubs of commercial R&D activity may fail to flow to industry and thus not benefit the economy. A second troubling implication is for the scientists themselves: two scientists of equal academic ability and with similar interests in having their work disseminated to the commercial world may be differently rewarded by virtue of working at an institution that is located inside vs. outside a hub of relevant R&D.

1. The Location of Academic Institutions and Knowledge Flow to Industry

A. The Flow of Academic Science to Industry

Scientific knowledge can increase R&D efficiency because it guides the invention process (Nelson 1982; Mokyr 2002). Large-scale empirical studies have established a link between university research and corporate patenting (Jaffe 1989) as well as productivity growth (Adams 1990). Using a survey of 77 major firms in various industries, Mansfield (1998) found that more than 5% ($44bn) of the total 1994 product sales of those firms were directly due to innovations that might not have been possible in the absence of academic research. Cohen, Nelson, and Walsh (2002) show that firms use academic knowledge both to generate new ideas and to address existing R&D problems.

Considering the economic value of academic discoveries, concerns have long been voiced that frictions might prevent the flow of that knowledge to industry. Charles Babbage argued that "the man of science should mix with the world" (Babbage 1832, 384), not only to ensure that he investigates important questions but also so that knowledge flows to manufacturers. Similarly, Mokyr (2002, 7)

proposes that "progress in exploiting the existing stock of knowledge will depend first and foremost on the efficiency and cost of *access* to knowledge." This paper explores this proposition empirically, focusing in particular on the impact of the geographic location of academic research institutions.

B. Geographic Location and the Flow of Academic Science to Industry

The flow of academic science to firms may be impeded by the fact that academic research institutions and industrial R&D labs are not always collocated. Audretsch and Feldman (1996) find that the bulk of innovative activity in the U.S. occurs on the coasts, especially in industries where scientific knowledge plays a decisive role. What complicates the process of knowledge diffusion is that academic research takes place in many locations, including at institutions not close to centers of commercial innovation. The sharp contrast between the geographic dispersion of academic scientists and the spatial concentration of industrial R&D activity is clearly visible in the case of the biotechnology industry. Analyzing biotechnology firms that completed an IPO in the early 1990s, Audretsch and Stephan (1996) link the location of biotechnology firms with that of academic scientists who had relationships with these companies. While 69% of the firms in their sample were based in Boston, the San Francisco Bay area, and San Diego, those regions accounted for only 36% of all academic contacts. A number of academic institutions elsewhere conducted leading-edge academic research but did not appear to be connected to any local biotechnology firms (e.g., Yale and the University of Texas Southwestern Medical Center).[2]

In principle, the distinct geographic distributions of academic scientists and the consumers of their discoveries in industrial R&D laboratories might not necessarily affect the flow of scientific knowledge. The academic environment is distinctive in its openness, and new discoveries are widely published (Merton 1973; Dasgupta and David 1994; Stephan 1996). Besides, location might not result in frictions if firms located further away from academic institutions do not have the same knowledge needs as those that are nearby. In practice, however, corporate inventors' limited visibility into the latest academic developments means that they might not exploit every potentially beneficial academic discovery. The flow of academic knowledge to corporate inventors might be imperfect because access to the scientific literature is not costless (Cohen and Levinthal 1989; Mokyr 2002). Inventors keep track of the latest academic developments in their fields through a variety of channels such as specialized journals and websites, professional conferences, through friends and colleagues, and by reading the work of academic scientists in their field. Considering inventors' limited time and resources, they are unlikely to be able to remain up-to-date with all relevant and useful academic knowledge in their field.

[2] In our data, public universities are generally located further from industry, perhaps due to land-grant provisions; however, controlling for private-vs-public institutions does not explain our findings.

The clearest evidence of frictions in the flow of knowledge may be the well-documented impact of geographic distance on the flow of knowledge. A number of studies have found that firms located in close proximity to academic institutions exploit their research more than firms that are located further away (Jaffe, Trajtenberg, and Henderson 1993; Zucker, Darby, and Brewer 1998; Furman and MacGarvie 2007; Belenzon and Schankerman 2013). This paper goes beyond prior work on geography and diffusion, which has focused largely on the distance separating an academic scientist and a particular firm that might utilize that scientist's knowledge. Instead, we focus on whether a focal academic scientist is located where similar commercial R&D is being conducted. In other words, the location of an academic research institution *inside or outside of a "hub" of relevant commercial R&D* will have important implications for knowledge flow, especially to distant firms. R&D "hubs" are central areas of knowledge generation and exchange among collaborators, competitors, and beyond. Commercial inventors trying to stay current with the latest innovations in their field may be more likely to be exposed to new academic developments that emerge in locations where similar commercial R&D is also happening.

Commercial inventors may be exposed to academic discoveries in the vicinity of such hubs for a variety of reasons. First, the concentration of R&D activity in a particular field may focus firms on that geographic region. While keeping up on developments of collaborators and competitors alike, they may thus become aware of academic discoveries in the same location. Second, because informal interactions between academic and commercial scientists are more likely to arise when the two groups are in close proximity (Mowery and Ziedonis 2015), commercial inventors in such hubs are likely to become aware of new discoveries by nearby academic scientists. To the extent that these commercial inventors then relay information regarding new discoveries to commercial inventors outside the hub, academic discoveries located near hubs of commercial R&D may flow to far-flung firms. Third, hubs of commercial R&D in particular fields may tend to host conferences and other formal gatherings of both commercial and academic scientists, further facilitating the flow of knowledge. By contrast, industry inventors may be more likely to overlook valuable discoveries that emerge outside those hubs.

2. Data Construction

A. Empirical Approach

Identifying the impact of the location of academic research institutions on knowledge flow to industry is nontrivial because the emergence of academic discoveries in specific locations is not exogenous to the geographic distribution of commercial R&D. 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 science. Because simultaneous discoveries constitute instances in which the same 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; Thompson and Fox-Kean 2005), these paper twins allow us to measure knowledge flow and non-flow 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) measuring the flow of academic discoveries to industry; and (3) the construction of "hubs" of commercial R&D that are relevant to a particular discovery.

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[3] (in this case: pages 481-490 and 491-501). As we detail in Appendix II, 46% of the simultaneous discoveries in our dataset correspond to back-to-back publications.

Figure I about here

We exploit simultaneous discoveries to overcome the aforementioned identification problem inherent to analyzing the impact of location on knowledge flow. In line with prior findings that the flow

[3] Editors sometimes decide to 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).

of knowledge is localized, we therefore consider that publication alone does not guarantee the perfect flow of academic knowledge to all inventors (Jaffe, Trajtenberg, and Henderson 1993; Azoulay, Graff Zivin, and Sampat 2012; Belenzon and Schankerman 2013). In practice, we measure the rate of dissemination by tracking the references to each scientific paper in patents. In the BID protein example, the paper located in Boston (where local firms perform R&D in similar fields) received more references from patents than did the paper in Dallas, which is largely isolated from relevant industry.

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). For convenience, its main principles are summarized in Appendix II. The full dataset of simultaneous discoveries consists of 1,246 papers and 578 simultaneous discoveries. From this set, we discard 50 papers published by firms as our aim is to study the flow of academic knowledge.[4] Given our interest in the flow of academic research to industry, we then drop 588 twin papers from simultaneous discoveries where none of the twin papers received any references from patents assigned to firms (see Appendix III). This could arise if none of the twin patents were referenced by any patent, or if they were referenced only by university patents (which we will later use in a placebo test). Finally, we removed 295 twin papers where any patent referencing one of the twin papers references all of the twin papers (and thus does not provide variation on our dependent variable; however, these are reintroduced in robustness). Excluding those leaves 313 twin papers reporting 146 simultaneous discoveries.

For each twin paper, we collect its geographic origin, the journal in which it was published, 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 well as papers in the top 15 scientific journals, the latter serving as a measure of the institution's prestige in the academic community. Summary statistics for all 1,196 academic twin papers are in Table I, segmented by whether none (588), all (295), or one (313) of the twins for a simultaneous discovery were referenced.

Table I about here

Table II provides a breakdown of the most frequent cities and institutions among the 313 twin papers in our analyses, for which one but not all twins for a simultaneous discovery were referenced. As

[4] Of the remaining papers, 43% are referenced by a firm-assigned patent. By comparison, Azoulay, Graff Zivin, and Sampat (2012) report that 12% of the academic publications are ever cited in patents. Our rate is likely higher because our sample is composed of particularly important discoveries.

is visible in Panel A, many of our twin papers are published at high-status institutions, including nearly 5% at Harvard University alone. This underscores both the importance of accounting for the prestige of the institution (since commercial inventors may be more likely to be exposed to such discoveries) as well as ensuring that our results are not driven by any one institution. 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 the Flow of Academic Science to Industry

Tracking the flow of academic science to industry is challenging because such flows can take a variety of forms including licensing, consulting, strategic partnerships (Roach and Cohen 2013). In a landmark paper, Jaffe, Trajtenberg, and Henderson (1993) proposed that patent citations can be used to measure knowledge flow. Patent citations are readily available yet have important limitations. First, patent citations have legal implications since they delimit the scope of an invention. 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), significantly complicating the task of using citations as a measure of knowledge flow. 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 studying the localization of knowledge flows (Thompson and Fox-Kean 2005; Henderson, Jaffe, and Trajtenberg 2005). Third, not all knowledge is observable by looking at patents (Griliches 1990); specifically, academic institutions primarily disclose knowledge through scientific publications rather than patenting, so their output is invisible through direct patent search (Agrawal and Henderson 2002; Belenzon and Schankerman 2013; Roach and Cohen 2013).

Measuring the flow of academic knowledge to industry via patent references to academic articles (e.g., Belenzon and Schankerman 2013; Azoulay, Graff Zivin, and Sampat 2012) helps overcome these challenges. Importantly, while patent-to-paper references cannot capture all relevant flows such as those occurring via private interactions,[5] 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). In addition, the large majority of simultaneous discoveries in our sample are in life sciences (to which most of the highest-impact-factor scientific journals belong); this is advantageous because in the

[5] The fact that references cannot measure this type of private and uncodified knowledge flow is likely to bias our results toward under-estimating the impact of location as a driver of knowledge flow because these private interactions tend to be more local (Mowery and Ziedonis 2015).

life sciences the use of publications and patents by firms is widespread, making scientific references *from* (but not *to*) patents a more accurate indicator of knowledge flow than they might be in other fields.

Another advantage of using patent-to-paper references stems from the fact that while every patent by definition represents a unique discovery; 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 can be published disclosing very similar knowledge (e.g., Cozzens 1989). Hence, it may be that an inventor is not aware of both papers reporting a single discovery that needs to be referenced by the patent as prior art.

Indeed, the U.S. Patent and Trademark Office imposes no duty on the inventor to reference every paper "twin" 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 one of them is sufficient. Of course the inventor may reference all relevant twin papers, and this is not uncommon as shown in Table I. Of the 608 twin papers where the simultaneous discovery was referenced by a patent, nearly half of the time (295) every twin was referenced whereas in the slight majority of cases (313) one twin but not all were referenced. That inventors do not always reference only one twin raises the question of whether they are unaware of the others vs. reluctant to reference them.

It seems unlikely that inventors would be reluctant to reference twin papers of which they are aware simply for reasons of cost or convenience. Many patents list dozens of scientific articles, and unlike some scientific publications the USPTO does not impose constraints as to the maximum number of references that can be included. Still, scientists tend to prefer citations to prominent peers in their scientific publications (Merton 1968), and one could be worried that the same might be true of inventors who might prefer references to publications stemming from more prestigious institutions. Our results regarding commercial R&D hubs facilitating the flow of knowledge from academia to industry might be questioned if inventors generally favored referencing papers by high-status authors or institutions that happened to be located in hubs. In addition, many prestigious institutions are located far from commercial hubs in specific fields. In line with this reasoning, we find no strong correlation between the academic prestige of a paper's author or institution and its likelihood of being referenced by a focal patent.

Unlike patent citations to other patents, references to scientific publications are not straightforward to analyze. Appendix III details our approach for linking papers and 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*.

D. Measuring the Location of Relevant Hubs of Commercial R&D

Establishing hubs of commercial R&D in a particular scientific field is not straightforward. Unlike national borders, state borders, or metropolitan statistical areas studied in prior work (e.g., Jaffe, Trajtenberg, and Henderson 1993; Thompson and Fox-Kean 2005; Singh and Marx 2013; Belenzon and Schankerman 2013), the location of such hubs are not readily available from administrative records. In fact, they are likely to be field specific and to evolve over time (Feldman and Florida 1994).

To measure whether an academic institution is located inside or outside a relevant hub of commercial R&D, we focus on inventive activity (a) in the relevant scientific field (b) within 5 years of the discovery and (c) within commuting distance of the institution. We label a location as a "hub" of commercial R&D for a USPTO subclass / 5-year period if more than 5% of all patents in that subclass/period are located within a 50-mile radius. The details of our algorithm and reasons for choosing these particular thresholds are given in Appendix IV.

To illustrate the concept of being located inside or outside of a hub of relevant commercial R&D, we return to our simultaneous discovery from Figure I. Again, we examine 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. In determining whether either of these research teams was inside or outside of the R&D hubs relevant to this simultaneous discovery, 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 have the same classification, yielding 17 unique subclasses. The next step is to locate "hubs" of commercial R&D in these technological areas. We find 3858 firm-owned patents that were assigned to these subclasses during 1995-1999. The locations with R&D "hubs" containing more than 5% of patenting activity for the above 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 locations. While Boston is within 50 miles of Canton, Massachusetts, Dallas is far from any of the cities listed. Thus we classify the paper published in Dallas as lying outside a hub of relevant commercial R&D. (Note: in the case where papers have authors from multiple institutions, we check whether any institution at which *any* of the authors is located is within commuting distance of the relevant R&D hubs.) Applying this definition, 79.9% of the papers found in our sets of twins are outside of a hub of relevant commercial

R&D. In many cases, both twin papers reporting a simultaneous discovery are located outside R&D hubs.

F. Empirical Setup

Our analysis leverages the simultaneous-discovery nature of our data since a patent that references one paper is presumably at a similar risk of referencing any of its "twins" as shown in Figure II. 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 references but also unrealized references to twins of a focal paper, we pair each patent that references one of the 313 papers with the other "twin" papers that disclose the same simultaneous discovery. That is, given a pair of twin papers 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, given that the twin papers disclose the same simultaneous discovery. This process yields 1,638 paper-patent dyads.

Figure II about here

For each paper-patent dyad representing a (potential) scientific reference, we account for both temporal and spatial separation between the paper and patent. Given that our key explanatory variable reports whether a paper is located outside a hub of relevant R&D activity, the distance between paper and potentially-referencing patent is perhaps our most important control. One might worry that papers lying outside of R&D hubs are simply further away from potentially-citing patents and thus may be less 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 twins reporting that discovery. Thus, for a given patent that is arguably at equal risk of referencing any of the twin papers, our analysis reveals the factors associated with a particular twin being referenced. The regression equation is given as

$$REFERENCED_{ijk} = f(\varepsilon_{ijk};\ \alpha_0 + \alpha_1 OUTSIDE\_R\&D\_HUBS_i + \alpha_{2\circ}\bar{X}_{ik} + \gamma_{jk})$$

where *j* represents the simultaneous discovery, *i* represents the paper reporting the simultaneous discovery, and *k* represents the potentially-referencing patent. $OUTSIDE\_R\&D\_HUBS_i$ is the main explanatory variable and is defined at the paper-patent dyad level as described in Appendix IV. 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 to one twin paper vs. another. 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 simultaneous discovery.

3. Empirical Results

A. Replication and Basic Results

We begin our analysis in Table IV. Before proceeding, we note the non-significance of most control variables in Table III (not shown). In particular, we find little correlation between academic prestige—as measured by the number of articles appearing in the top 15 scientific journals—and the likelihood of a particular paper being referenced. Although it might seem that inventors would be more likely to reference papers from more prominent authors and institutions, this does not appear to be the case in our sample. To some extent this may reflect our selection of twin papers, which may be generally well-known; in a broader sample of papers from a wider variety of journals, institutional prestige may play more of a role. Author characteristics and journal impact factor are likewise uncorrelated with the likelihood of being referenced. For our particular sample it does not appear that authors are strategically preferring one twin paper over another due to the status of the paper's author, journal, or institution.

In column (1) we replicate prior findings regarding the spatial separation of the focal paper and a potentially-referencing patent. Consistent with work on the geographic localization of knowledge diffusion, distance is negatively associated with the likelihood of a paper being referenced by a patent. That our analysis of simultaneous discoveries yields similar localization dynamics to those seen in prior diffusion studies helps to allay concerns that this set of 313 papers might exhibit strongly different characteristics than larger samples analyzed previously.

Table IV about here

In column (2) of Table IV, we include an indicator of all authors of a focal paper located outside "hubs" of relevant commercial R&D. Adding this covariate does not materially affect the estimate of the coefficient on the distance control from column (1), but its own coefficient is negative with statistical significance at the 1% level. Paper twins located outside of R&D hubs are 9.97% less likely to be referenced than their twin(s) located inside R&D hubs. Thus it appears that the deleterious effect of being separated from a potential user of academic knowledge can be ameliorated if the knowledge producer is collocated with communities of commercial inventors in the same field. To return to our earlier example, although we might expect a firm to be less likely to build upon an academic discovery 1,000 miles away when an equivalent discovery is only 100 miles away, the more distant discovery may in fact be more likely to be referenced if it is within a hub of commercial R&D in that field.

Following Singh and Marx (2013), in column (3) we switch from parametrically modeling distance to a non-parametric series of dummies in order to capture the nuances of localization. Compared to an omitted category of more than 2,500 miles separating the patent and paper, a reference is 20.4%

more likely to occur when the patent and paper are within 20 miles of each other. This sharp dropoff with even a small separation between patent and paper resembles the Belenzon and Schankerman (2013) finding that the probability of a paper being referenced by a patent drops by 40% when they are within 25 miles. The coefficient on location outside an R&D hub is largely unaffected by this change.

The evaluation of new academic knowledge requires specialized skills, so the effect of being located in a R&D hub should therefore only be salient when those hubs are specific to the discovery. As a placebo test of the field-specificity of this mechanism, in column (4) we replace our hub definition from Appendix IV with a more general measure of biotech clusters as defined by Hoffman and Fucht (2014). Although we do not purposefully focus on the life sciences, 14 of the top 15 scientific journals are in this area. Replacing our field-specific hub definitions with these broader biotech cluster definitions does not yield statistically significant results.

In column (5) we perform another placebo test. The analyses in Table IV to this point measure references to academic papers from patents assigned to firms in order to measure the flow of knowledge from academia to industry. If commercial R&D hubs affect only academia-to-industry flows and not knowledge diffusion more generally, the geography of commercial R&D should not affect flows *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. While there is a negative relationship between the likelihood of an academic paper being referenced by a university patent and its location outside a hub of relevant commercial R&D, the coefficient is very imprecisely estimated. Taken as a whole, Table IV indicates that hubs of commercial R&D facilitate the flow of academic knowledge to industry, not within academia itself, and only when the hubs are highly specific to the science at hand.

B. Robustness

In Table V, we test the robustness of the relationship between an academic paper being referenced by a firm-assigned patent when it is located in a hub of relevant R&D. The concentration of twin papers in certain cities or at particular institutions, as suggested by Table II, may raise the possibility that our finding is driven by a few key locations. Columns 1-3 report one of a series of leave-one-out tests to address this concern. Boston, Massachusetts is home to 8.3% of our twin papers, yet the relationship between location in a hub and referencing is robust to its omission in column (1). Similar results are also recovered when sequentially excluding the next top five cities: San Diego, CA; New York, NY; Bethesda, MD; and San Francisco, CA. In column (2), we repeat the leave-one-out test for academic institutions with a high percentage of papers. Harvard is host to 4.8% of our 313 twin papers; omitting either it or any of five other top institutions yields similar results. Our result also does not depend on any one assignee, as shown in column (3): Senomyx Inc. is the most frequent, with 5.4% of patents.

In column (4) we test our definition of a hub of commercial activity as having at least 5% of

patenting activity in that specific field. It is quite unusual to find locations with a fifth or a quarter of patenting, so we do not test 20% or 25% thresholds, but when we double the percentage of patenting in that field required to 10% we find results are consistent with those obtained at the 5% level.

Our analyses thus far have focused on the 313 twin papers reporting simultaneous discoveries where one (but not all) twins were referenced by a patent, as this provides variation in the dependent variable. In the remaining columns of Table V we adopt alternative specification in order to include additional twin papers. Unlike the conditional logit, the linear probability set-up in column (5) includes the 295 twin papers reporting simultaneous discoveries where all twin papers were referenced by every patent that references any of them. Our result is preserved, with approximately a 9% lower probability of papers located outside of hubs being referenced by industry patents.

In column (6) we analyze all 1,196 academic twin papers, including the 588 papers reporting simultaneous discoveries where none of the twin papers were referenced by industry patents. Given that we cannot construct patent-paper dyads for the 588 papers where no patent references them or their twins, instead we adopt just the twin paper as our unit of observation. Our dependent variable also changes to be the count of references to a given twin paper from any patent. Correspondingly, we adopt a negative binomial specification as indicated by a goodness-of-fit test following Poisson regression. Consistent with our prior models, twin papers located outside of relevant hubs of commercial R&D are less likely to receive references from industry patents. We do not employ simultaneous-discovery fixed effects in column (6) as the 588 papers reporting simultaneous discoveries where no twins were referenced would be dropped for lack of variation in the dependent variable. However, re-estimating the count model using OLS enables inclusion of the simultaneous-discovery fixed effects and yields similar results (unreported).

C. Interaction effects and mechanisms

In Table VI we further explore the mechanisms underlying the attenuation of knowledge flow to industry from academics located outside of R&D hubs. Given our hypothesis that locating within commercial R&D hubs will facilitate awareness by a larger set of industry inventors, we expect that our results will be attenuated by factors that compensate for non-hub location by promoting awareness of the focal paper.

Our first moderator examines the relationship between the institution and industry. If proximity to relevant R&D hubs facilitates the flow of knowledge from academia to industry, then more formal relationships between academia and industry may substitute for the informal interactions that may arise given proximity to a hub. We explore this interaction in column (1a) by introducing a measure of commercial investment in R&D at the institution where a given paper was published. (Given that this data is collected from the Association of University Technology Managers, our analysis in column (1a) is necessarily limited to North American institutions.) Although the positive coefficient on the interaction of this measure and the indicator for being outside of relevant R&D hubs is suggestive of a substitution

effect, it is somewhat imprecisely estimated. In column (1b) of Panel A, we replace the interaction term between R&D hubs and commercial investment in R&D at a focal institution with a set of indicators interacting whether the paper is outside an R&D hub with four levels of commercial funding of R&D at the institution: (1) no funding; (2) funding totaling less than $5MM; (3) more than $5MM but less than $22MM, i.e. the 75th percentile; (4) more than $22M. (In column (1b), as well as (2) and (3), the omitted category is all papers inside an R&D hub.) The estimated coefficients on the interaction variables suggest that the apparent substitution effect in column (1a) is driven primarily by papers written at institutions that are outside R&D hubs and which do not receive any commercial funding.

Table VI about here

Given the challenges of interpreting interaction terms in nonlinear models such as conditional logit, we repeat this analysis with a linear probability model and graph the resulting coefficients in Figure III. Panel A corresponds to column (1b). The only estimated coefficient significantly different from zero is for papers outside of relevant R&D hubs which receive no R&D investment dollars from industry. Thus it appears that the negative effect of being located outside of R&D hubs is felt most acutely along the margin of institutions that lack formal connections to industry.

Figure III about here

In Panel B of Table VI, we consider whether institutional reputation or prestige might generate exposure for papers located outside of R&D hubs. As noted earlier, academic prestige does not generally affect the probability of a twin paper being referenced. However, it might do so for the subset of papers published outside of R&D hubs as prestige could compensate for those papers' lower visibility among commercial inventors. We interact the outside-R&D-hub indicator with indicators for the four quartiles of institutional prestige and rely again on the graphed linear-probability coefficients for inference. Panel B of Figure III suggests that papers whose institutions are in the lowest quartile of prestige are those that are the most clearly affected by their location outside of a commercial R&D hub.

Finally, we examine the interplay between separation from R&D hubs and the distance between the paper and the potentially-referencing patent. As in the previous two panels, in Panel C of Table VI we add interaction terms for all distance dummies, including for the previously-omitted category of more than 2500 miles separating the two. The pattern revealed by 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 within commuting distance of each other (i.e., less than 20 miles, or 20-50 miles), being isolated from R&D hubs appears not to have an effect. Given the finding from Table IV that paper-patent dyads within 20 miles of each other are much more likely to contain a reference, it appears that the paper being outside an R&D hub may not matter if it is close enough to the potentially-citing patent. This is consistent with a model in which academic researchers and commercial inventors each have strong

internal linkages within their respective communities but weak connections to each other. The resulting cross-community gap can apparently be bridged by collocation, either with the inventors on the possibly-referencing patent, or with the community of inventors in an R&D hub for that specific scientific field.

By contrast, when the separation between the focal paper and focal patent is more than 2500 miles, hubs appear to lose their efficacy in facilitating the flow of knowledge from academia to industry. One might be concerned that this reflects the influence of bicoastal networks between say, San Diego and Boston, which facilitate distant knowledge flow independent of R&D hubs. However, only 18.5% of paper-patent dyads separated by more than 2500 miles are within the U.S., while all but one of the non-U.S. dyads with similarly large separation are on different continents. Both for the intercontinental and within-U.S. dyads separated by more than 2500 miles, the fraction of papers referenced by a focal patent is lower than for dyads separated by fewer than 2500 miles. Thus it appears that there are limits to the ability of hubs to facilitate the flow of knowledge across continents.

In sum, the analyses of Table VI and Figure III suggest that papers located outside hubs of relevant R&D are most disadvantaged when their institutions lack formal connections with industry, when the institution has low prestige, and when the paper and potentially-citing patent are neither within commuting distance nor extremely separated.

One possible mechanism includes linkages between the inventor on a focal patent and those inventors in the commercial hubs of R&D that a focal paper is near. If inventors in those hubs become aware of nearby academic discoveries through local interactions with academic scientists, awareness of those discoveries may flow more broadly within industry via social networks. As described in Appendix V, we calculate the network overlap between the hub(s) associated with a particular paper and the inventors of a possibly-referencing patent. In unreported results, a patent is considerably more likely to reference a paper if there exists network overlap between itself and the inventors in the paper's hub(s), if any. However, such overlap does not strongly mediate the effect of hubs. This could be either because networks play only a small role in the flow of academic science to industry, or because the networks observable to us represent a small subset of interpersonal interactions along which such flows occur.

4. Discussion

This paper proposes a methodology to identify the impact of the geographic location of academic institutions on the flow of scientific knowledge to industry. We use simultaneous discoveries—i.e., events in which multiple teams of scientists share credit for a discovery to identify a new factor underlying how the location of academic research institutions impacts the flow of scientific knowledge to firms. As corporate inventors attempt to keep track of the newest academic findings, they not only privilege local discoveries but are also particularly exposed to knowledge produced in the relevant hubs of commercial R&D. This effect is attenuated when the inventors on a focal patent becomes exposed to the paper by

other means including the prestige of the academic institution, its formal connections to industry, or its collocation with those same inventors.

We interpret our results cautiously, for several reasons. Even though simultaneous discoveries allow us to examine the same knowledge emerging at different locations, the non-referencing of an academic paper in a corporate patent is not a perfect measure of the absence of knowledge flow since those discoveries could conceivably diffuse in ways that do show up in the patent process. In addition, our set of twin papers is relatively small and largely concentrated in the life sciences (albeit not by design), limiting generalizability to other industries. Although we do not observe period effects in our data, it is possible that in a broader set of simultaneous discoveries additional heterogeneity may emerge (perhaps also with regard to prominence of the journal, author, or institution).

Despite these limitations, our findings raise a number of questions. First, the current distribution of academic research organization might promote equal access to science across geographical areas, but our results suggest that such efforts toward egalitarianism may come at a cost by complicating firms' exploitation of discoveries made at remote academic institutions. This raises something of a dilemma for science policy. Should policy-makers abstain from funding academic research conducted outside of the relevant R&D hubs, or should they instead promote the dissemination of scientific knowledge produced at those institutions? Similarly, what are the steps that those institutions may take to offset their inherent disadvantage due to location? Finally, are the careers of the scientists who accept positions at those institutions negatively affected by their reduced ability to have an impact beyond academia?

Firms may also be able to take advantage of the fact that valuable scientific discoveries emerging from institutions located outside R&D hubs tend to be ignored. Not only might broadening technological search reveal otherwise-missed opportunities; less competition for such discoveries at institutions collocated with neither the focal firm nor hubs of commercial R&D might yield attractive licensing terms.

Our analysis points to the importance of uncovering the costs and benefits of the organization of academic science. Most empirical studies of this question use large-scale citation analysis with a difference-in-difference approach (Murray and Stern 2007; Agrawal and Goldfarb 2008; Furman and Stern 2011). Exogenous shocks, however, are not available for every important question. This paper attempts to enrich the "empiricist's toolbox" by describing a new approach exploiting the occurrence of simultaneous discoveries in science. Our hope is that this study will contribute to a better understanding of academic science as an institution both by offering theoretical insights about the impact of academic location and by establishing the value of simultaneous discoveries as a research methodology.

Appendix I:
Cross-sectional Analysis

A naïve approach to estimate the association between the geographic location of academic research institutions and knowledge flow to industry would be to consider references in corporate patents to a large sample of academic publications. We examined all 28,133 academic papers published in the top 15 scientific journals between 2000 and 2010. These are *Nature, Science, Cell, the New England Journal of Medicine, Journal of the American Medical Association, Lancet, CA: Cancer Journal for Clinicians, Nature Genetics, Nature Materials, Nature Medicine, Nature Immunology, Nature Nanotechnology, Nature Biotechnology, Cancer Cell, and Cancer Stem Cell*.

To determine whether these academic papers are located outside "hubs" of relevant industrial R&D, as detailed in Appendix IV we need to know which USPTO patent subclasses are relevant to the focal paper,[6] which are available only for papers that receive one or more references from patents as described in Appendix III. Hence, this analysis is conditional on having received a reference from a patent assigned either to a university or a firm. There are 1,649 such papers among the 28,133, or 5.9%.

For those papers, we are able to assess whether they emerged in a relevant corporate R&D hub by observing the distribution of corporate patents from the relevant subfield in the 5 years surrounding the publication of the paper. To assess knowledge flow to industry, we count how many times those 1,649 publications are referenced by patents assigned to *firms* (not universities). A simple difference-of-means test indicates that the average number of references for such papers located outside relevant hubs of industry R&D is 1.7 as compared with 3.2 for papers located inside a hub of relevant commercial R&D, with statistical significance at the 0.01% level. Similar results are recovered in unreported regression models that incorporate the controls from Table I.

[6] In defining hubs for cross-sectional analysis, we face the following tradeoff: since we cannot establish the field of a publication that is not referenced in any patent, we can either define corporate hubs broadly by building a measure that depends on corporate patent density but is not field specific or we can sacrifice sample size and focus instead on those academic publications that receive at least one patent reference. Both approaches lead to the result that academic publications emerging inside corporate R&D hubs receive significantly more patent references.

Appendix II:
An Automated Method to Build a List of Simultaneous Discoveries

The algorithm is rooted in the results from two distinct literatures. On the one hand, sociologists of science have found that citations provide a window into the scientific community's allocation of credit. In a sense, the community uses citations as a "vote" regarding which team deserves the credit for a given discovery (Cozzens 1989). As a result, systematic co-citation in the scientific literature indicates that the community has decided that the credit for a specific discovery ought to be shared across different teams. While occasional co-citation might point to discoveries that are complementary rather than simultaneous, systematic co-citation indicates that two of more papers share the credit for the same discovery. On the other hand, citations provide a convenient similarity metric to relate documents (Marshakova 1973; Small 1973). As such, they can be used to map science, but can also be fed into search engines pointing to related papers. As an example, as CiteSeer uses co-citations to compute the relatedness between academic papers (Giles, Bollacker, and Lawrence 1998). Recent studies have suggested that these algorithms can be made even more precise by considering citation proximity within each paper. For instance, papers that are co-cited in the same sentence tend to be particularly similar to each other (Gipp and Beel 2009; Tran et al. 2009). The algorithm that was used here goes one step further and considers pairs of scientific publications that are consistently cited together—i.e., in the same parenthesis, or adjacently.

In practice, the algorithm uses five steps. In step 1, a dataset consisting of information about 42,106 scientific articles was built using ISI Web of Knowledge. It is composed of all the non-review research publications that appeared in the 15 scientific journals having the highest impact factor between 2000 and 2010. In step 2, each reference in all of these articles were given a unique identifier using Pubmed and CrossRef. Of 1,294,357 references, 744,583 unique references were identified. Step 3 generates a database of pairs of all references that were (a) co-cited at least once, (b) written no more than a calendar year apart, (c) have no overlapping authors, (d) in which at least 5 citations for each reference are observed in the dataset of 42,106 citing articles. Of the 17,050,914 pairs of papers that were considered, 449,417 pairs meet these criteria. Step 4, consists in establishing a first measure of co-citation. A Jaccard co-citation coefficient was used following the scientometric literature. It consists in the intersection over the union of citations that both papers receive for each pair. 2,320 pairs of papers were selected that had a co-citation coefficient superior to 50%. Finally, step 5 consists in selecting those pairs for which 100% of the co-citations took place in the same parenthesis or adjacently. To do so, a parsing algorithm examined all the co-citing articles. 495 pairs for which fewer than 3 co-citing articles could be parsed were excluded. Of the remaining 1,825 pairs, 720 had been cited adjacently in 100% of the co-citing articles. These 720 pairings of 1,246 papers disclose 578 unique discoveries since there are instances of discoveries involving three or more teams.

The extent to which the resulting pairs are actually instances of simultaneous discoveries was tested in several ways. First, if they really are twins, our pairs of scientific papers should be published around the same time. The algorithm matches on co-citation and not on publication month. If two alleged paper twins were not really disclosing the same discovery, one would expect them to be on average six months apart or more.[7] The 720 paper twins in the entire dataset were in fact published on average 1.8 months apart, a lag considerably shorter than the average time between paper submission and publication. In fact, 373 pairs of twins were published the exact same month, and 267 of them were published in the same issue of the same journal. Second, the Pubmed related citation algorithm uses semantic similarity to match scientific papers. Since the large majority of the 1,246 papers also appear in Pubmed, we can use this algorithm to measure the semantic similarity between pairs of papers that our algorithm identified as disclosing the same discovery. If the pairs were not very closely related, they should not be using the same words and should therefore be ranked far from each other. Pubmed ranks two papers of the same pair right next to each other 42% of the time. The rank difference is inferior to 10 for 90% of the pairs. Third, 27 scientists who had been corresponding authors on at least one of the 1,246 papers were interviewed. Importantly, none of them contested the fact that they were sharing the credit with another team for the same discovery and some were bitter about it.[8] Five of the interviewees claimed that their idea had been stolen by the other team. Confirming that the algorithm uses very conservative criteria, the interviewees also revealed in several cases that more teams than we were aware of had claimed to have taken part in the simultaneous discovery. One should keep in mind that, by design, our algorithm excludes any priority claim that is not clearly visible through the citations of the broader scientific community.

---

[7] The algorithm does not match on month, but it limits the consideration set of papers to pairs that were published no more than a calendar year apart (we considered that papers published more than 23 months apart cannot be disclosing the same discovery). This choice is limiting because many independent discoveries are known to have taken place years apart of each other (see Ogburn and Thomas (1922) for numerous examples). However, since credit for scientific discoveries is a function of priority, it is reassuring that we ended up with pairs of papers published very close to each other. Besides, for our study, it is important that the paper emerge around the same time so they have the same chance of being used by corporate inventors.

[8] Sharing the credit does not mean that the two (or more) papers were identical. Two scientific articles written by two different teams are never completely identical, and differences might exist in the tools/methods used, in the number of experiments, or in the interpretation of the results. However, the fact that the papers share the credit indicates that the scientific community considers that both teams provided convincing evidence to support their claim of priority in making the discovery.

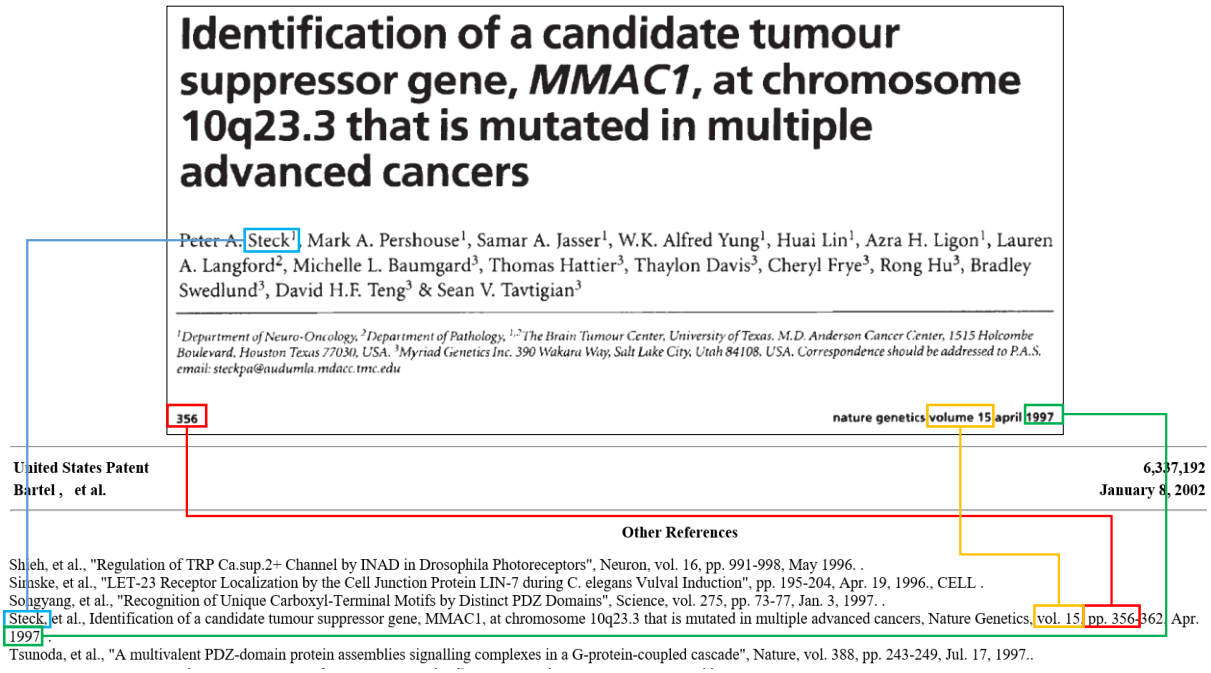
Appendix III:
Capturing References from Patents to Papers

Tracking references to papers is more difficult than references to patents. As seen in Figure A.1, papers are listed as free-text strings. One might match on title and journal name but from our initial attempts to do so we found frequent abbreviations of title and journal names as well as occasional misspellings.

Instead, we elected to use four more reliably matched criteria: 1) the surname of the first author, 2) the year of article publication, 3) the volume number of journal, and 4) the starting page number. This tuple is highly unlikely to be non-unique; in order for this to occur, two authors with the same surname would have had to publish articles in different journals that had the same volume number in the same year; moreover both articles would have to start on the same page.

We automatically parse the first author's name, year, volume, and first page from the scientific references listed in the patent. These fields are also extracted from the scientific papers, for which this data is available in a more structured format. The two groups of {author surname, year of journal, journal volume, initial page number} characteristics are matched with each other. We use the matches produced from these four criteria as a first pass to create a superset of possible matches and then inspect those by hand for Type II errors (less than 2% of automated matches were dropped as false positives).

Figure A.1: References to Patents vs. Papers



Notes: The paper and patent above illustrate the process of finding scientific references. Instead of attempting to match on title or journal name, only the first author's name, year, volume number, and initial page number are used. In some patents referencing this article, the journal name is abbreviated in various ways (*Nat. Genet.*; *Nature Genets.*; etc.). In others, the article title is omitted, such as in patent 6,287,854 where the reference appears as *Steck et al (Apr. 1997) Nature Genetics 15, 356-362.*

Appendix IV:
Constructing "Hubs" of Commercial R&D in Specific Fields

This measure is operationalized as follows. We start by collecting the technological subclassifications from all patents, whether industry or academic, that contain references to one of the 313 "twin" papers 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 categorized into 712 unique subclasses. For each subclass, we then collect all non-university patents belonging to that subclass, whether or not they reference any of the twins in our study. We find a total of 1,430,822 corporate patents that were categorized by the USPTO into one of the 712 technology subclasses.

We then construct "hubs" of commercial R&D activity as follows. For each of the 712 technology subclasses that characterize our simultaneous discoveries, we collect the locations in which those non-university patents are found in that subclass. For each location, we count the number of patents in that same subclass within a 50-mile radius for each half-decade. We divide those two figures to yield the percentage of overall patenting activity from that technology subclass occurring in that location. We label a location as a "hub" of R&D for that subclass if more than 5% of patents in that technology subclass are located within a 50-mile radius. 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), we require that a location have at least five patents in that subclass to qualify as a "hub." This exercise yields a list of R&D hubs for each of the 712 technology subclasses relevant to our simultaneous discoveries within five years of the publication date. (Some subclasses are widely distributed across locations and thus do not have any hubs.)

To determine whether a given academic paper is inside or outside of a relevant hub of industrial R&D, we first make a list of the technological subclasses for all patents that referenced either the focal paper or any of its twins. These patent subclasses delimit the relevant scope of R&D activity for that simultaneous discovery. For each twin paper reporting that simultaneous discovery, we then check whether there is at least one R&D hub within 50 miles (i.e., commuting distance) of the institutional affiliation of *any* author on the paper. It is important to note that location with regard to relevant R&D hubs is a paper-level attribute, neither an institution- nor city-level attribute. Institutions and cities may be inside a hub for one field but outside of R&D hubs for others. For example, in the 1995-1999 period, Dallas is not considered a biotechnology R&D hub but it is a hub for semiconductor R&D; the opposite is true for Boston. It is also possible that the concentration of R&D shifts over time, which motivates our use of five-year windows for determining hubs.

Appendix V:
Mapping network overlap between a focal patent and paper hubs

Our objective is to detect interpersonal linkages between a focal patent and a focal paper via which information regarding the paper might flow to the owner of a patent. Of course, a full inventory of all such interactions is unobservable, and mapping networks across domains (i.e., from academia to industry) is nontrivial. As a proxy, we utilize information regarding patent inventors to construct second-degree network connections. For each inventor on any patent in our dataset, we assemble the list of that inventor's co-inventors on that or any other patent (i.e., that inventor's first-degree connections). We then find the list of the co-inventors for that inventor's co-inventors (i.e., that inventor's second-degree connections).

Our initial approach is to detect first- or second-degree overlap between the corresponding author of the academic paper and the inventors of the patent. As approximately one-fourth of the authors of the 313 twin papers in our sample ever filed a patent, by definition this mapping is limited to those authors. For a given author of a paper (who has at least one patent) and a potentially-citing patent, we check whether any of the author's first-or-second-degree connections is also a first-or-second-degree connection of any of the inventors on the focal patent. For the approximately one-quarter of paper authors who have a patent, we find zero instances of overlap between the paper's author and the inventors on the focal, possibly-referencing patent in the dyad. Note that this does not mean there is no network overlap, only that we cannot detect such using patent records. Directly mapping the names of paper authors as well as their collaborators, students, advisors, etc. to patent holders' names may further illuminate the nature of these network connections.

Our second approach is to locate connections between the inventors on a focal patent and the inventors in relevant hubs of commercial R&D for a given paper. Again, such hubs may facilitate the flow of information from academia to industry when inventors in those hubs are linked to the inventors of potentially-citing patents. For each academic paper located inside one or more hubs, we gather the inventors of all commercial patents defining the hub and then assemble their first- and second-degree co-inventors. We then check for overlap between these connections and those of the inventors on a focal patent that might reference the focal paper.

Using this second method, we find that 9% of paper-patent observations where the paper is inside a hub of commercial R&D contain a network overlap between the inventors on the focal patent and the inventors in the hub. Again, we do not claim to have captured all network overlap but only that which is detectable using patent data. Some paper-patent combinations have up to nine overlapping first-and-second-degree connections between the focal patent and the patents in the hub.

References

Adams, James D. 1990. “Fundamental Stocks of Knowledge and Productivity Growth.” *Journal of Political Economy* 98 (4): 673–702.

———. 2002. “Comparative Localization of Academic and Industrial Spillovers.” *Journal of Economic Geography* 2 (3): 253–78.

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

Agrawal, Ajay, and Avi Goldfarb. 2008. “Restructuring Research: Communication Costs and the Democratization of University Innovation.” *The American Economic Review* 98 (4): 1578.

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. doi: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. doi:16/j.respol.2008.12.001.

Audretsch, David B., and Maryann P. Feldman. 1996. “R&D Spillovers and the Geography of Innovation and Production.” *The American Economic Review* 86 (3): 630–40. doi:10.2307/2118216.

Audretsch, David B., and Paula E. Stephan. 1996. “Company-Scientist Locational Links: The Case of Biotechnology.” *The American Economic Review* 86 (3): 641–52.

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*, 107–55. University of Chicago Press. http://www.nber.org/papers/w16683.

Babbage, Charles. 1832. *On the Economy of Machinery and Manufactures ... Second Edition Enlarged*. Charles Knight.

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

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

Cohen, Wesley M., and Daniel A. Levinthal. 1989. “Innovation and Learning: The Two Faces of R & D.” *The Economic Journal* 99 (397): 569–96. doi:10.2307/2233763.

Cohen, Wesley M., Richard R. Nelson, and John P. 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.

Dasgupta, Partha, and Paul A. David. 1994. “Toward a New Economics of Science.” *Research Policy* 23 (5): 487–521.

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.

Feldman, Maryann P., and Richard Florida. 1994. “The Geographic Sources of Innovation: Technological Infrastructure and Product Innovation in the United States.” *Annals of the Association of American Geographers* 84 (2): 210–29. doi:10.1111/j.1467-8306.1994.tb01735.x.

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. doi:10.1016/j.jebo.2006.05.014.

Furman, Jeffrey, and Scott Stern. 2011. “Climbing atop the Shoulders of Giants: The Impact of Institutions on Cumulative Research.” *American Economic Review* 101 (5): 1933–63.
Galasso, Alberto, and Mark Schankerman. 2014. “Patents and Cumulative Innovation: Causal Evidence from the Courts*.” *The Quarterly Journal of Economics*, November, qju029. doi:10.1093/qje/qju029.
Giles, C. L, K. D Bollacker, and S. Lawrence. 1998. “CiteSeer: An Automatic Citation Indexing System.” In *Proceedings of the Third ACM Conference on Digital Libraries*, 89–98.
Gipp, B., and J. Beel. 2009. “Citation Proximity Analysis (CPA)-A New Approach for Identifying Related Work Based on Co-Citation Analysis.” In *Proceedings of the 12th International Conference on Scientometrics and Informetrics (ISSI’09)*, 571–75.
Griliches, Zvi. 1990. “Patent Statistics as Economic Indicators: A Survey.” *Journal of Economic Literature* 28 (4): 1661–1707. doi:10.2307/2727442.
———. 1998. “Introduction to ‘R&D and Productivity: The Econometric Evidence.’” In *R&D and Productivity: The Econometric Evidence*, 1–14. University of Chicago Press. http://www.nber.org/books/gril98-1.
Grossman, Gene M., and Elhanan Helpman. 1993. *Innovation and Growth in the Global Economy*. MIT Press.
Halim, Nadia. 2000. “Bridging Apoptotic Signaling Gaps.” *The Scientist*, August 21. http://www.the-scientist.com/?articles.view/articleNo/12974/title/Bridging-Apoptotic-Signaling-Gaps/.
Harris, Gardiner. 2011. “Federal Research Center Will Help Develop Medicines.” *The New York Times*, January 22, sec. Health / Money & Policy. http://www.nytimes.com/2011/01/23/health/policy/23drug.html.
Henderson, Rebecca M., Adam B. Jaffe, and Manuel Trajtenberg. 2005. “Patent Citations and the Geography of Knowledge Spillovers: A Reassessment: Comment.” *The American Economic Review* 95 (1): 461–64.
Hoffman, William, and Leo Fucht. 2014. *The Biologist’s Imagination: Innovation in the Biosciences*. Oxford University Press.
Jaffe, Adam B. 1989. “Real Effects of Academic Research.” *The American Economic Review* 79 (5): 957–70.
Jaffe, Adam B., Manuel Trajtenberg, and Rebecca M. Henderson. 1993. “Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations.” *The Quarterly Journal of Economics* 108 (3): 577–98. doi:10.2307/2118401.
Lampe, Ryan. 2012. “Strategic Citation.” *Review of Economics and Statistics* 94 (1): 320–33. doi:10.1162/REST_a_00159.
Lemley, Mark A. 2007. “Should Patent Infringement Require Proof of Copying?” *Michigan Law Review* 105 (7): 1525–36.
Mansfield, Edwin. 1998. “Academic Research and Industrial Innovation: An Update of Empirical Findings.” *Research Policy* 26 (7-8): 773–76. doi:10.1016/S0048-7333(97)00043-7.
Marshakova, I. V. 1973. “System of Document Connections Based on References.” *Nauchno-Teknicheskaia Informatsiia* 2 (1): 3–8.
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.
———. 1968. “The Matthew Effect in Science The Reward and Communication Systems of Science Are Considered.” *Science* 159 (3810): 56.
———. 1973. *The Sociology of Science: Theoretical and Empirical Investigations*. University of Chicago Press.
Mokyr, Joel. 2002. *The Gifts of Athena: Historical Origins of the Knowledge Economy*. Princeton University Press.
Mowery, David C., and Arvids A. Ziedonis. 2015. “Markets versus Spillovers in Outflows of University Research.” *Research Policy* 44 (1): 50–66. doi:10.1016/j.respol.2014.07.019.
Murray, Fiona. 2002. “Innovation as Co-Evolution of Scientific and Technological Networks: Exploring

Tissue Engineering." *Research Policy* 31 (8-9): 1389–1403.
Murray, Fiona, and Scott Stern. 2007. "Do Formal Intellectual Property Rights Hinder the Free Flow of Scientific Knowledge?: An Empirical Test of the Anti-Commons Hypothesis." *Journal of Economic Behavior & Organization* 63 (4): 648–87. doi:10.1016/j.jebo.2006.05.017.
Nelson, Richard R. 1959. "The Simple Economics of Basic Scientific Research." *Journal of Political Economy* 67 (3): 297–306.
Nelson, Richard R. 1982. "The Role of Knowledge in R&D Efficiency." *The Quarterly Journal of Economics*, The Quarterly Journal of Economics, 97 (3): 453–70.
Niehans, Jurg. 1995. "Multiple Discoveries in Economic Theory." *European Journal of the History of Economic Thought* 2 (1): 1. doi:Article.
Ogburn, William F., and Dorothy Thomas. 1922. "Are Inventions Inevitable? A Note on Social Evolution." *Political Science Quarterly* 37 (1): 83–98.
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. doi:10.1287/mnsc.1120.1644.
Romer, Paul M. 1990. "Endogenous Technological Change." *Journal of Political Economy* 98 (5 pt 2). http://www.dklevine.com/archive/refs42135.pdf.
Singh, Jasjit, and Matt Marx. 2013. "Geographic Constraints on Knowledge Spillovers: Political Borders vs. Spatial Proximity." *Management Science* 59 (9): 2056–78. doi:10.1287/mnsc.1120.1700.
Small, Henry. 1973. "Co-Citation in the Scientific Literature: A New Measure of the Relationship Between Two Documents." *Journal of the American Society for Information Science* 24 (4): 265–69.
Sohn, Eunhee. 2014. "The Endogeneity of Academic Science to Local Industrial R&D." *Academy of Management Proceedings* 2014 (1): 11413. doi:10.5465/AMBPP.2014.286.
Stephan, Paula E. 1996. "The Economics of Science." *Journal of Economic Literature* 34 (3): 1199–1235.
Stigler, George J. 1980. "Merton on Multiples, Denied and Affirmed." *Transactions of the New York Academy of Sciences* 39 (1 Series II): 143–46. doi:10.1111/j.2164-0947.1980.tb02774.x.
Thompson, Peter. 2006. "Patent Citations and the Geography of Knowledge Spillovers: Evidence from Inventor- and Examiner-Added Citations." *Review of Economics and Statistics* 88 (2): 383–88. doi:10.1162/rest.88.2.383.
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.
Tran, Nam, Pedro Alves, Shuangge Ma, and Michael Krauthammer. 2009. "Enriching PubMed Related Article Search with Sentence Level Co-Citations" 2009: 650–54.
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. doi:10.1257/jep.25.2.199.
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=1,196.

		Mean	Median	Std. Dev.	Min.	Max.
Academic "twin" papers from simultaneous discoveries where no paper was referenced (N=588)	Number of patents referencing twin paper	0.000	0	0.000	0	0
	Journal impact factor	3.003	3.280	0.630	0	3.959
	Paper located in U.S.	0.581	1	0.494	0	1
	Paper was patented	0.102	0	0.303	0	1
	Corresponding author stock of patents	0.995	0	3.622	0	55
	Corresponding author stock of papers	73.997	41	87.133	0	754
	Institution's 5-year stock of patents	185.688	35	341.872	0	2308
	Institutional prestige	95.802	30	147.989	0	577
Academic "twin" papers from simultaneous discoveries where every twin was referenced (N=295)	Number of patents referencing twin paper	2.692	1	5.009	0	38
	Paper authors outside hubs of relevant R&D	0.902	1	0.298	0	1
	Journal impact factor	3.023	3.467	0.666	0	3.959
	Paper located in U.S.	0.651	1	0.478	0	1
	Paper was patented	0.149	0	0.357	0	1
	Corresponding author stock of patents	1.312	0	3.935	0	41
	Corresponding author stock of papers	73.634	40	89.547	0	679
	Institution's 5-year stock of patents	153.264	60	239.953	0	1606
	Institutional prestige	99.853	42	129.426	0	577
Academic "twin" papers from simultaneous discoveries where one but not all twins were referenced (N=313)	Number of patents referencing twin paper	8.518	5	12.034	0	81
	Paper authors outside hubs of relevant R&D	0.799	1	0.402	0	1
	Journal impact factor	3.062	3.467	0.632	0	3.959
	Paper located in U.S.	0.645	1	0.479	0	1
	Paper was patented	0.214	0	0.411	0	1
	Corresponding author stock of patents	1.652	0	5.437	0	75
	Corresponding author stock of papers	76.047	44	93.427	0	447
	Institution's 5-year stock of patents	144.336	63	221.796	0	1415
	Institutional prestige	101.667	38	141.625	0	577

Notes: The construction of "twin" papers reporting simultaneous discoveries is detailed in Appendix II.

Table II
Location of 313 academic "twin" papers where one but not all twins 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 as described in Appendix II. The subset of 313 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 and correlations for paper-patent dyads (N=1,638)

	Mean	Median	Std. Dev	Min	Max
Twin paper referenced by focal patent	0.477	0	0.5	0	1
Paper authors outside hubs of relevant R&D	0.667	1	0.472	0	1
Paper outside biotech clusters	0.524	1	0.5	0	1
Journal impact factor	3.202	3.51	0.581	0	3.959
Paper located in U.S.	0.664	1	0.472	0	1
Paper was patented	0.24	0	0.427	0	1
Corresponding author stock of patents	0.439	0	0.738	0	4.331
Corresponding author stock of papers	3.577	3.688	1.326	0	6.463
Institution's 5-year stock of patents	3.181	4.04	2.314	0	7.256
Institutional prestige	3.36	3.82	2.011	0	6.36
Publication lag, paper vs. patent	5.142	4	3.236	0	17
Distance between paper and patent	7.167	7.747	1.952	0	9.263
Paper and patent in same state	0.096	0	0.294	0	1
Paper and patent in same country	0.51	1	0.5	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 discoveries being referenced by industry patents

	Dependent variable indicates that the "twin" paper was referenced by a patent assigned to a				
	firm				**university**
	(1)	(2)	(3)	(4)	(5)
Paper authors outside hubs of relevant R&D		-0.767**	-0.791**		-0.380
		(0.297)	(0.289)		(0.287)
Paper outside biotech clusters				-0.291	
				(0.294)	
Distance between paper and patent	-0.182*	-0.187**			
	(0.0709)	(0.0709)			
Paper and patent <20 miles apart			1.851*	1.398+	-0.923
			(0.798)	(0.785)	(1.038)
Paper and patent 20-50 miles apart			0.452	0.103	-0.654
			(0.775)	(0.784)	(1.042)
Paper and patent 50-250 miles apart			0.872	0.498	-0.480
			(0.651)	(0.682)	(0.404)
Paper and patent 250-1000 miles apart			0.731+	0.662+	-1.122***
			(0.385)	(0.388)	(0.335)
Paper and patent 1000-2500 miles apart			0.348	0.300	-0.874*
			(0.343)	(0.348)	(0.359)
Paper and patent in same state			-0.380	0.0311	-0.00643
			(0.493)	(0.489)	(0.746)
Paper and patent in same country			-0.237	-0.0690	1.724**
			(0.573)	(0.579)	(0.546)
Observations	1,638	1,638	1,638	1,638	1,071
# twin articles	313	313	313	313	378
Pseudo-R2	0.122	0.143	0.147	0.133	0.0946
Log-likelihood	-503.3	-491.8	-489.1	-497.2	-339.1
Simultaneous-discovery/patent FE	YES	YES	YES	YES	YES

Notes: Observations are academic-paper/firm-assigned-patent dyads. All models are estimated using conditional logit and include simultaneous-discovery/patent fixed effects. All models include controls for the paper (U.S.-based, journal impact factor, discovery was patented), author (stock of patents and papers), and institution (stock of patents and papers) characteristics as well as characteristics of the paper-patent dyad (publication lag). Papers outside hubs of relevant R&D are 9.97% less likely to be referenced. Standard errors are clustered at the level of the simultaneous discovery; *** p<0.01; ** p<0.1; * p<.0.05; + p<0.10.

Table V
Robustness checks

	conditional logit				linear probability	negative binomial
	leave-one out tests			10% hub threshold		
	omit top city	omit top institution	omit top assignee			
	(1)	(2)	(3)	(4)	(5)	(6)
Paper authors outside hubs of relevant R&D	-0.818*	-0.705*	-0.893**	-1.027*	-0.0870*	-1.564***
	(0.393)	(0.303)	(0.295)	(0.454)	(0.0356)	-0.175
Constant					-0.133	1.523***
					(0.326)	(0.0580)
Observations	1,182	1,371	1,565	1,638	5,138	1,196
# twin articles	252	286	311	313	608	1,196
(Pseudo) R-squared	0.137	0.107	0.178	0.146	0.038	0.0291
Log-likelihood	-450.7	-487.3	-489.9	-357.3		
Simultaneous-discovery/patent FE	YES	YES	YES	YES	YES	NO

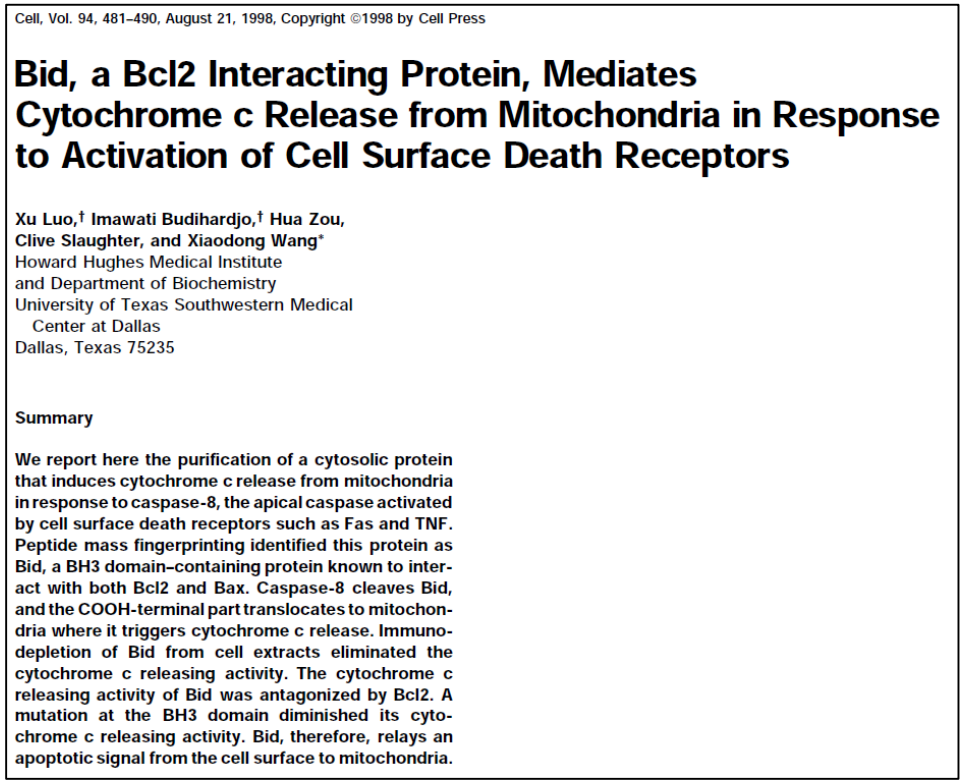
Notes: For columns (1-5), observations are academic-paper/firm-assigned-patent dyads; the dependent variable indicates whether the patent in the dyad references the paper. Column (5) employs a linear probability model, which enables estimating the model using the 295 twin papers where every patent referencing one twin also referenced all other twins. In column (6), observations are all 1,196 academic twin papers; the dependent variable counts the number of patents referencing a focal paper. (Overdispersion indicates a negative binomial.) All models include controls for characteristics of the paper (U.S.-based, journal impact factor, discovery was patented), author (stock of patents and papers), and institution (stock of patents and papers). Columns (1-5) also control for characteristics of the paper-patent dyad (publication lag; spatial distance). Standard errors are clustered throughout at the level of the simultaneous discovery; *** p<0.01; ** p<0.1; * p<.0.05; + p<0.10.

Table VI
Interaction effects

Panel A Industry investment in the focal institution	(1a)	(1b)	**Panel B** Institutional prestige	(2)	**Panel C** Distance between focal paper and patent	(3)
Paper authors outside hub of relevant R&D	-1.290**					
	(0.449)					
Industry $ funding research at institution	-0.0336+	-0.662				
	(0.0184)	(0.564)				
Outside hubs * industry $ funding institution	0.0296+	0.0630	Outside hubs, lowest quartile	-1.925**	Outside hubs, within 20m	-0.516
	(0.0170)	(1.507)		(0.646)		(1.146)
Outside hubs, no industry funding		-2.408**	Outside hubs, second-lowest quartile	-0.726	Outside hubs, 20-50m	-0.449
		(0.808)		(0.745)		(1.565)
Outside hubs, little industry funding		-1.421	Outside hubs, second-highest quartile	-1.032*	Outside hubs, 50-250m	-1.972+
		(1.462)		(0.445)		(1.111)
Outside hubs, more industry funding		-0.866+	Outside hubs, highest quartile	-0.622	Outside hubs, 250-1000m	-2.372**
		(0.482)		(0.401)		(0.805)
Outside hubs, most industry funding		-0.455			Outside hubs, 1000-2500m	-1.691**
		(1.067)				(0.626)
					Outside hubs, >2500m	-0.173
						(0.341)
Observations	874	874		1,638		1,638
# twin articles	162	162		313		313
Pseudo-R2	0.204	0.242		0.164		0.173
Log-likelihood	-242.8	-231.1		-479.4		-474.3
Simultaneous-discovery/patent FE	YES	YES		YES		YES

Notes: Observations are academic-paper/firm-assigned-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), author (stock of patents and papers), 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 *inside* 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 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 I
Example of "twin" papers reporting a simultaneous discovery



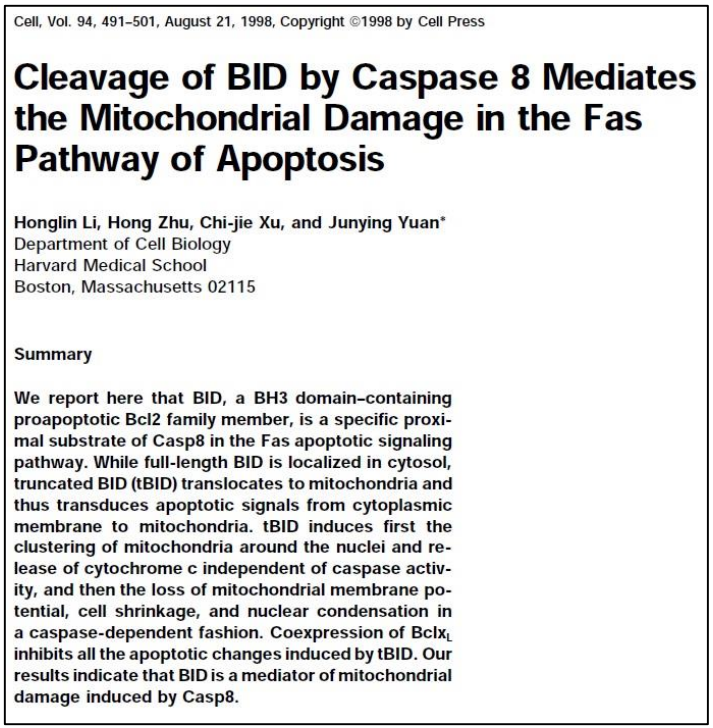
Cell, Vol. 94, 481–490, August 21, 1998, Copyright ©1998 by Cell Press

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

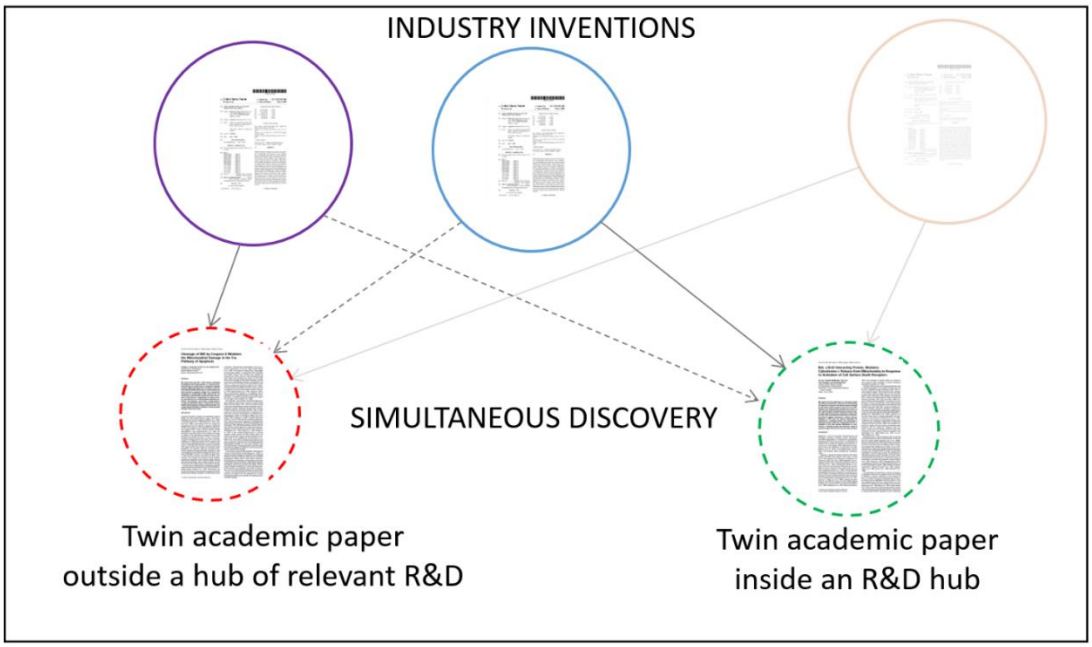
Cleavage of BID by Caspase 8 Mediates the Mitochondrial Damage in the Fas Pathway of Apoptosis

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.

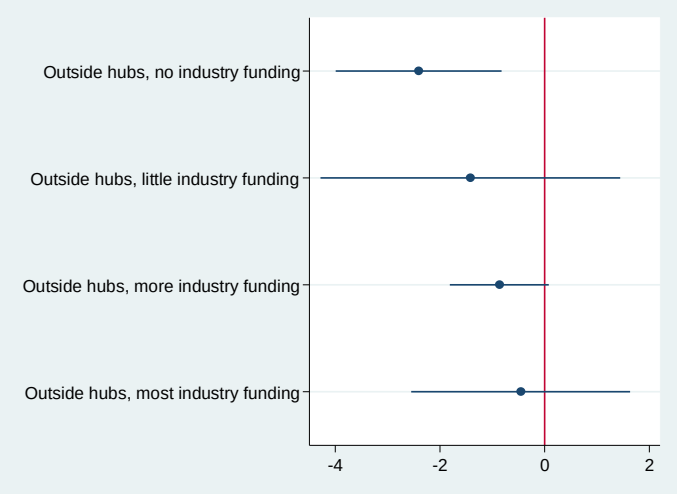
Figure II
Construction of paper-patent dyads



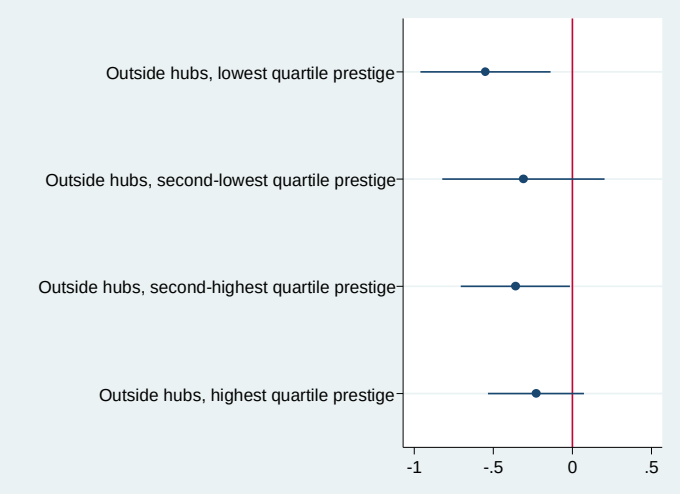
Notes: The figure depicts two "twin" papers reporting a simultaneous scientific discovery. Three patents reference one or both of the papers, as represented by solid arrows. Each of these realized patent-to-paper references constitutes an observation. In addition, dotted lines represent possible but unrealized patent-to-paper references in that the other "twin" paper reporting the same simultaneous discovery could reasonably have been referenced by the same patent. Note that the patent referencing both twin papers is dimmed as the two observations represented by its solid arrows provided no variation in the dependent variable and are thus excluded from our conditional logit estimation. However, results are robust to a linear-probability specification which includes the dimmed observations.

Figure III
Interaction effects for papers outside of relevant R&D hubs with other factors.

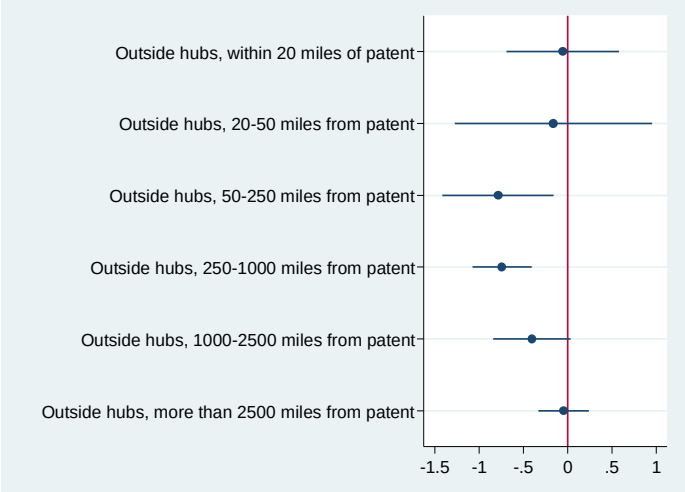
Panel A: Funding of R&D at the paper's institution by industry



Panel B: Organizational prestige, defined as # of papers in the top 15 scientific journals



Panel C: Distance between paper and patent



Notes: Coefficients are plotted from linear probability models with the identical setup as Table VI.