Cross-regional Ties within Firms:

Promoting Knowledge Flow or Discouraging Knowledge Spillover?

Minyuan Zhao
Stephen M. Ross School of Business
University of Michigan
701 Tappan St., E4616
Ann Arbor, MI 48109
(734) 647-6978
myzhao@umich.edu

Mazhar Islam
Carlson School of Management
University of Minnesota
321 19th Avenue S, Room 3-365
Minneapolis, MN 55455
(612) 624-3582
mislam@csom.umn.edu

We thank Gautam Ahuja, Sendil Ethiraj, Jeff Macher, Marie Louise Mors, Rachelle Sampson, Myles Shaver, and Bernard Yeung for valuable comments and suggestions.

Cross-regional Ties within Firms:

Promoting Knowledge Flow or Discouraging Knowledge Spillover?

ABSTRACT

R&D activities are increasingly carried out by collaborators from different geographic locations. Such collaborations are expected to promote knowledge flow across distance and generate positive spillover to the local community. However, little attention has been paid to the role of firm organization surrounding these collaborations. While cross-regional collaborations in entrepreneurial firms usually involves extensive interpersonal interactions, those in large, established organizations are often highly structured and routinized, which may even increase internal interdependence and raise barriers to cross-organizational learning. Examining the pharmaceutical industry from 1975 to 2001, we find that collaborations play an important role in bridging the locally clustered R&D activities. Nevertheless, cross-regional ties in the big pharma companies contribute significantly less – sometimes even reduce knowledge spillover – to local innovation, and the pattern persists over time. Interviews conducted in several R&D clusters corroborated the empirical findings.

Keywords: R&D, knowledge spillover, internalization

JEL codes: F23, O32, O33

I. INTRODUCTION

In January 2004, the Swiss giant Roche announced the opening of its fifth global pharmaceutical R&D center in Shanghai, China. The entry of Roche is widely expected to give a boost to the quality of local R&D, especially to the R&D of new ventures recently set up by Chinese scientists and returnees. Meanwhile, the company emphasized that "the new center will be part of Roche's global pharmaceutical R&D network." In particular, the local employees will be working with their colleagues in Penzberg (Germany) on anti-cancer medicines, in Palo Alto (U.S.) for anti-HIV targets, in Basel (Switzerland) on therapies for the Alzheimer disease, and in Nutley (U.S.) for new treatment of obesity. How will such cross-regional collaborations affect the expected knowledge spillover to local firms?

Two literatures, focusing on two different aspects of the phenomenon, seem to provide opposing answers to this question. The knowledge transfer literature argues that collaboration is an effective means of knowledge transfer, both within and cross firms (Cockburn & Henderson, 1998; Fleming, *et al.*, 2004; Zucker & Darby, 2005; Singh, 2005a). Through joint problem solving (McEvily & Marcus, 2005), geographically separated team members can gain access to the tacit knowledge that is otherwise locally bounded (Jaffe, *et al.*, 1993; Audretsch & Feldman, 1996; Szulanski, 1996), hence bringing fresh perspectives to the local community.

The innovation organization literature, however, is taking a different view. In established organizations, the reliance on formal structures and operational routines reduces the need for interpersonal exchanges (Scott & Davis, 2006: 38-40). Assuming pre-specified roles in a collaborative relationship, individual researchers may find it difficult to convey information to outsiders due to their partial understanding of the overall technology structure (Rajan & Zingales, 2001). Meanwhile, without the complementary knowledge embedded in a faraway location, local firms may see little incentive in learning from their innovative neighbors (Zhao, 2006).

This paper aims to reconcile the two literatures by introducing organizational heterogeneity into our studies of knowledge flow and knowledge spillover. The specific phenomenon of cross-regional R&D collaboration provides an interesting case for this purpose. First, there has been a large literature on the roles of interpersonal ties (Reagans & Zuckerman, 2001; Fleming, *et al.*, 2004) as well as geographic proximity (Jaffe, *et al.*, 1993; Almeida & Kogut, 1999) in knowledge transfer, so we can easily benchmark our results against the established findings. Second, in most studies on knowledge transfer, firms are either the unit of analysis or used as a control variable for studies on interpersonal networks. However, both the formation and the functioning of interpersonal networks may be actively shaped by the organizational environments around them. Despite the large literature on collaborative R&D, the interaction between organizational heterogeneity and collaborative ties is still understudied.

We hypothesize that the effects of cross-regional collaborations on localized knowledge spillover would vary with the types of firm organization that the collaborations are situated in. While connections with the outside world are beneficial to local R&D in general, those organized in large established firms would also increase the internal complementarity and interdependence among the firms' geographically dispersed units, thus raising barriers to cross-organizational learning.

The hypothesis is tested with data from the pharmaceutical industry between 1975 and 2001. In this period, the birth of contemporary technologies created "a burst of new companies" (Chandler, 2005). Meanwhile, the fast development of information technologies has also made it easier for researchers to access distant knowledge and engage in cross-regional collaborations. Such features offer a rich setting for us to assess the role of firm organization in knowledge spillover and innovation.

The empirical results show that cross-regional collaborations play an important role in bridging the locally clustered R&D activities. Such collaborations not only increase the value of innovations directly resulting from the collaborations, but also benefit innovations in the whole community. However, the contribution of cross-regional collaborations formed by the largest pharmaceutical companies is found to

be significantly lower than that of similar collaborations formed by their smaller counterparts, and the gap remains large over time. This suggests that the development of information technologies in the past decades, while facilitating knowledge flow across distance, may have also intensified the internal linkages and promoted strategic R&D organization within established firms, thus discouraging knowledge spillover across firm boundaries. The findings are further corroborated by our extensive interviews in some the large pharmaceutical clusters. Interestingly, the local researchers reported much more helpful communications with the "well-connected" but "simple" firms in the community, than with the large global-oriented R&D labs that "feel like isolated islands."

The rest of the paper is organized as follows. The next section develops the theoretical framework and the main hypotheses. Section III introduces the data sources and describes the empirical models used in the analysis. The empirical results and robustness checks are discussed in Section IV and Section V, respectively. Section VI concludes.

II. THEORY DEVELOPMENT

In this section, we first discuss the role of cross-regional collaborations in technology transfer across distance and knowledge spillover to the local communities. Then, we incorporate firm heterogeneity into the discussion and examine how organizational contexts would affect the nature of interpersonal linkages. The theoretical concepts derived in this section lead to the empirical setup that follows.

2.1 Localized spillover and cross-regional collaboration

Innovations often result from the combination and recombination of existing knowledge (Schumpeter, 1939; Fleming, 2001). In the pharmaceutical industry, for example, drug discovery usually requires the inputs of scientists skilled in a very wide range of disciplines (Henderson & Cockburn, 1994). As R&D projects become larger and more complex over time, collaborations among scientists become increasingly indispensable in the innovation process (Arora & Gambardella, 1994; Jones, 2005). In fact, the

percentage of U.S. patents produced by teams of three or more inventors steadily increased from 16.4% in 1975 to 34.8% in 1995. For pharmaceutical patents, the numbers are 25.7% and 43.7%, respectively.

Meanwhile, different geographic regions remain specialized in different knowledge bases (Cantwell & Janne, 1999; Verspagen & Schoenmakers, 2004). Even within the pharmaceutical industry, we observe the concentration of biotechnology firms in Boston and the dominance of cardiovascular equipment manufacturers in Minneapolis. Calculating the technological distance (Jaffe, 1986) between each pair of metropolitan areas in the U.S., based on both pharmaceutical patents and patents in all categories, we find little sign of technological convergence across regions in the past three decades.

With geographically bounded knowledge pools, scientists from different locations may find it beneficial to collaborate with one another. Technology transfer has proved to be challenging even within the same firm (Szulanski, 1996). Joint problem-solving arrangements facilitate the acquisition of tacit and complex knowledge by providing a forum for experimentation, observation, and search for solutions (McEvily & Marcus, 2005). Thus, cross-regional collaborations allow researchers to pool together a richer knowledge set, and their personal interactions help alleviate the difficulty of knowledge transfer across geographic distance (Lahiri, 2003; Frost & Zhou, 2005; Singh, 2005b).

The impact of cross-regional collaborations, however, is not limited to the firms that are organizing the collaborations. Such long distance linkages also bring fresh perspectives and expand the horizon for all neighboring firms through localized knowledge spillover. Since firms are more likely to search for and apply knowledge around their own technological positions (Cohen & Levinthal, 1990; Stuart & Podolny, 1995), collocation of similar firms can promotes cross learning in the local community. Nevertheless, too much localized connections may also stop firms from identifying new trends or exploring novel ideas, partly due to the recirculation of redundant information (Uzzi & Spiro, 2004). March (1991) argues that exploiting certainties at the expense of exploring new possibilities can be detrimental to a firm in the long run. Similar argument should also apply to geographic clusters: A region with little interaction with the

outside world is less likely to remain dynamic and innovative. In this sense, technologies representing cross-regional collaborations can be great sources of knowledge for local innovators.

2.2 Firm organization and knowledge internalization

The phenomenon of cross-regional collaborations is interesting by itself, but we cannot fully understand the role of such collaborations without taking firm organization into consideration. Collaborations occur in various institutional contexts, among university scientists, individual researchers, partners in small start-ups, as well as engineers of large multinational companies. Hence, the nature of knowledge transfer among the collaborators is also shaped by the organizational structure they are embedded in.

The interpersonal interactions involved in R&D collaborations are cited as the main reason why collaborations promote the diffusion of intellectual capital (Zucker & Darby, 2005). In entrepreneurial firms when collaborations are less structured, researchers often engage in extensive experimentation and explorations in order to achieve certain goals (McEvily & Marcus, 2005), which leads to comprehensive understanding of the problem they are facing. Mutual trust and deep appreciation for each other's work are crucial for a collaborative relationship to be productive and sustainable. In the interviews conducted in Shanghai's Zhangjiang Hi-tech Park, many researchers working at the local branches of U.S. startups acknowledge that they collaborate extensively with their American colleagues, who are sometimes their graduate school classmates. "We talk a lot, about literally everything, and we keep each other updated," a researcher said.

This may not be the case in a more formal organizational structure, where the role of each individual is well appointed and team members form "stable expectations" regarding the behaviors of other members, independently of their personal attributes (Simon, 1997). With well-established routines (Nelson & Winter, 1982; Cohen, 2006), collaborations can be achieved without much interaction among the individual team members. Each person is simply doing his or her part of the job, and their intellectual products are integrated through the hierarchical organizational structure. This is indeed the impression

we had when visiting some of the large multinational R&D labs: A good team member simply means "doing what he/she is supposed to do." Often times, coordination is taken care of by a team leader or manager, not by the loosely structured, but intensive communications among team members. When asked whom they would turn to for information, many answered "the boss," even though they admit that their foreign colleagues "are the experts on this."

In recent years, large established corporations have been playing an important role behind the ever-increasing occurrences of cross-regional collaborations. Multinational firms have long been recognized for their capacity to assimilate, generate and integrate knowledge on a global basis (Bartlett & Ghoshal, 1990; Feinberg & Gupta, 2004). Long-established institutions and internal deployment of employees across subunits help facilitate collaborations that would not be possible otherwise. The development of information technologies further strengthen the coordination abilities of the multi-unit, multi-location firms and allow them to spread out their R&D activities worldwide (Alcacer, *et al.*, 2005). Because of the complex internal organizations that large firms have, their internal long-distance ties may not have the same spillover effect as those established by smaller ventures.

From the capability point of view, with greater combinative capabilities (Kogut & Zander, 1992), large multinational firms are usually advantageous at internalizing their R&D and appropriating value from new technologies (Buckley & Casson, 1976). The large number of elements in an organization and the complex interactions among them also increase the difficulty of imitation (Rivkin, 2000). Cross-regional collaborations, if carried out through organizational routines, only strengthen the firm specificity of these interactions and limit effective communications across firm boundaries (Levin, 1988).

Furthermore, from the knowledge protection point of view, large firms may strategically organize their R&D activities in order to discourage learning by competitors. When a firm's R&D network spans multiple locations, at each location it can develop technologies that closely relate to the firms' internal resources residing elsewhere around the world. Since specialized and co-specialized complementary

assets are critical to the successful commercialization of an innovation (Teece, 1986; Anand & Galetovic, 2004), firms can minimize information outflow by increasing the reliance of local capabilities on knowledge and resources not readily available in the neighboring community. For example, in areas where intellectual property protection is weak, firms tend to intensify the monitoring of local R&D and make sure that only certain stages of the discovery process are carried out locally; the resulting technologies are quickly integrated into the firm's global knowledge base (Zhao, 2006). If cross-regional collaborations inside the firm serve as means of building internal complementarity, then the benefits gained from such collaborations are less likely to be shared by the local community.

III. DATA DESCRIPTION AND EMPIRICAL SETUP

Based on the theoretical discussion in the previous section, this section sets up the empirical frameworks to examine whether the role of cross-regional collaboration varies across firm organizations. In the following subsections, we first describe the data sources, the key variables, and a set of background statistics for this study. Then, two alternative econometric models, which address the same question from two different angles, are presented.

3.1 The pharmaceutical industry

We choose the pharmaceutical industry for this study for a number of reasons. First, it is one of the most knowledge intensive industries. On average, pharmaceutical firms spend about 20% of their revenue on R&D, and innovation is directly associated with firm performance. Therefore, how to benefit from knowledge flow and how to prevent the leakage of proprietary information become crucial questions for firms in this industry.

Second, pharmaceutical R&D has been highly concentrated geographically. A century after the industry pioneers such as DuPont and Parke-Davis established the first industrial R&D facilities around major research universities (MacGarvie & Furman, 2005), the majority of pharmaceutical R&D is still

conducted in the largest technology clusters: the New York-New Jersey-Philadelphia region, greater San Francisco, London, Rhine Valley, etc.

Third, this is a truly global industry and the players in the industry are extremely heterogeneous. Big pharmas such as Pfizer and GSK each employs thousands of researchers worldwide, while a typical biotech startup consists of only 2–3 scientists. Firm heterogeneity allows us to disentangle the role of firm organization from other factors in knowledge spillover.

Finally, with the pharmaceutical industry, we can take advantage of the rich information from the patent data (Henderson & Cockburn, 1996; Penner-Hahn & Shaver, 2005), and use the physical addresses of patent inventors to track the firms' R&D activities. Both the Yale Survey (Levin *et al.*, 1987) and the Carnegie Mellon Survey (Cohen *et al.*, 2000) found that patents play an especially important role in protecting intellectual capital of firms in the pharmaceutical industry. In addition, using patents granted by U.S. Patent and Trademark Office (USPTO) for the study of global R&D is justified by the special status of the U.S. pharmaceutical industry. According to recent *IMS Health* reports, the U.S. market accounts for nearly half of pharmaceutical sales worldwide, and U.S. companies control over 60 percent of the global pharmaceutical market. It is therefore reasonable to assume that most of the important pharmaceutical innovations would be filed for U.S. patent protection.

We apply a broad definition of the industry. Following Hall *et al.* (2001), the pharmaceutical patents used in this study include the following USPTO primary patent classes: 424, 514 (drugs), 128, 600, 601, 602, 604, 606, 607 (surgery & medical instruments), 435, 800 (biotechnology), and 351, 433, 623 (miscellaneous – drug & medical). We choose patent classes instead of SIC codes for industry classification because the industry involves many well diversified firms, and not all information on SIC codes is available. Dummy variables will be used to account for the possible variations across the 14 patent classes that are not captured by other independent variables.

To emphasize on firms' internal organization, we also exclude patents with multiple assignees, which are more likely to represent joint ventures or strategic alliances. Patents assigned to individuals, universities and other non-profit organizations are also excluded from the sample, although their inclusion does not seem to make any significant difference to the results. The final sample contains 204,139 patents[1], of which 129,071 – or 62.5% of the total – have at least one American inventor. These patents were applied between 1975 and 2001, and were granted before the end of 2004.

3.2 The geography of collaboration and knowledge spillover

To study cross-regional collaborations, we first have to define the "regions." In the benchmark analysis, a region is defined as a "metropolitan statistical area" in the U.S. – according to the U.S. Bureau of the Census – or a country in the rest of the world. For the pharmaceutical industry, we prefer to use metropolitan areas rather than states because some important technology clusters span multiple states (e.g., New York-New Jersey-Connecticut tri-state area) and some states contain multiple clusters (e.g., California). For the foreign countries, few have multiple pharmaceutical R&D centers in one country, and even if they do, imposing more refined definition throughout the world may introduce more noise than insightful information. The definitions of "county" or "prefecture," for example, vary widely across countries.

This definition generates 361 unique regions, including 263 metropolitan areas led by New York-New Jersey, San Francisco and Boston, and 98 foreign countries led by Japan, Germany and U.K. Among these regions, the top five percentile regions are associated with more than 63% of the patents in the sample. For robustness checks, we also use (1) country, (2) states, and (3) economic areas defined by the Bureau of Economic Analysis, as alternative definitions of regions. The advantage of economic areas, compared with metropolitan areas, is that they encompass both rural and urban counties.

[1] The number of observations in some regressions may be larger that this number because the same patent may be observed at multiple locations.

The emergence of new markets and the unprecedented development of information technologies over the past several decades have had profound influence on the way R&D is carried out worldwide. Figure 1 depicts the geographic distribution of R&D activities for the largest pharmaceutical companies between 1975 and 2001. For each firm and each year, we count the number of locations the firm has inventors in, and calculate the Herfindahl Index of its geographic concentration based on inventor locations. When the index is close to 1, the firm is concentrating almost all its R&D in the central lab, most likely at headquarters. Not surprisingly, the large firms significantly increased their geographic diversification of R&D activities during this period.

Meanwhile, as shown in Figure 2, the average distance among patent collaborators has been unambiguously increasing.

Insert Figures 1 and 2 here

However, these do not necessarily indicate that distance is disappearing. It may well be the case that geographic proximity is getting more important with the fast movement of knowledge frontiers (Leamer & Storper, 2001; Sonn & Storper, 2004). As knowledge spillover is still highly localized, researchers need to stay in the clustered areas in order to keep up with the technological frontier. As a result, firms that hope to access the localized knowledge have to be present in multiple locations, and rely on long-distance teams for technology integration.

What we observe in Figures 3 and 4 seem to support this argument. Here we measure the geographic distances between all the citation dyads in the pharmaceutical industry, and plot them along the citation years. Interestingly, the within-firm citations and cross-firm citations exhibit starkly different trends. While the citation distances inside the firms increased sharply between 1975 and 2001 – a trend consistent with the organizational changes – the citation distance across firms actually decreased. That is, knowledge spillover across organizational boundaries has become increasingly localized in this period!

Insert Figures 3 and 4 here

The above figures provide some background information on cross-regional collaborations. Next, we set up the empirical framework to examine the spillover effect of such collaborations on the overall innovation activities in the local community, and how this effect varies with firm organization.

3.3 Empirical setup

Patent is used as the unit of analysis, as it can capture interesting variations at the technology, firm and location levels. Two alternative models are used for the empirical analysis. Model I examines the *process* of knowledge spillover, where the focal patents are the sources of spillover. That is, we count the patent citations from the neighboring innovations to a focal patent, and associate the number of local citations with the focal patent's cross-regional collaboration behavior. Model II examines the *result* of knowledge spillover, where the focal patents are the beneficiaries. That is, we measure the overall quality of local patents, and associate it with the cross-regional collaborations observed in the region. Essentially, from two different perspectives, the two models test the same questions: Are cross-regional collaborations good for local innovation, and how does firm organization affect this relationship?

3.3.1 Model I: local citations to the focal patent

Patent citations are believed to be highly correlated with actual and perceived knowledge spillover (Jaffe, *et al.*, 2000; Jaffe *et al*., 2001; Lahiri, 2003), despite the considerable noise contained in this measure (Alcacer & Gittelman, 2005). In Model I, the dependent variable *local_cite* is defined as the number of citations from local innovations to the focal patents, excluding self-citations. To capture the spillover to local small ventures, we also calculate *local_small_cite*, a subset of *local_cite* that only includes citations from firms with less than ten patents during the observation year.

The extent of knowledge spillover generated by a focal patent is affected by two key factors. The first is *cross_region*, a dummy variable indicating whether the focal patent is developed by inventors from different geographic locations. The second is *big_pharma*, a dummy variable indicating whether the patent belongs to a firm whose pharmaceutical patent output is among the top five percentile of the whole industry during the observation year. It turns out that these firms filed around half of the patents in the sample. Alternative definitions of *big_pharma* – such as the top 50 pharmaceutical companies in terms of global sales – are used for robustness tests.

There are two reasons why we prefer to use the *big_pharma* dummy instead of a continuous measure of firm size. First, from the organizational point of view, there are some non-linear, qualitative differences between the world's largest pharmaceutical "empires" and the other large firms. Second, using a dummy variable can help us illustrate the marginal effects more effectively. Admittedly, this is an imperfect proxy for the concept of large established organizations, even though long heritage and hierarchical organization characterize most of the companies on the big pharma list. (See the table in Appendix.) In the regressions, we will also utilize the continuous measure to verify the robustness of the results.

What interests us the most, however, is the interaction between *cross_region* and *big_pharma*, i.e., whether the impact of cross-regional collaborations varies across different firm organizations. Put it differently, do we observe systematic differences between collaborations organized in large established firms and those in other organizational contexts?

A series of patent, firm and regional characteristics are applied as control variables. For example, we control for the number of scientists on the team (*inventors*), which may indicate the importance of the project and the budget associated with it. Intuitively, a patent developed by a large organization may obtain more future citations because of higher visibility, and a patent developed in a densely populated technology center should expect more citations due to localized knowledge flows. Hence, we count the

total number of pharmaceutical patents owned by the firm (*firm_size*) and the number developed in the region (*region_size*) for every observation year, and use their natural logarithms as control variables.

An important characterization of multi-location firms is their relative specialization at each location. For instance, some firms may prefer to apply the same expertise worldwide, while others exercise well-designed division of labor among their internal units. To capture this variation across firms, we first calculate the overall technological similarity between every pair of regions, following Jaffe (1986):

$$s_{ij} = \frac{v_i v_j'}{\sqrt{(v_i v_i')(v_j v_j')}} \in (0,1) \qquad (1)$$

where v_i and v_j are two vectors representing two regions, and the k^{th} element of v_i is the number of patents developed in region i that fall into the k^{th} patent class. A nice feature of this measure is that the absolute number of patents does not matter; only their structural distribution does. $s_{ij} = 1$ when the two vectors exactly overlap and $s_{ij} = 0$ when they are orthogonal. Next, we go through the same calculation for each firm, and take the difference between the firm-level measure and the overall s_{ij} for each regional pair (i, j). Aggregating the results at the firm level generates *loc_similarity*, the variable that measure whether a firm's innovation is more homogeneous (less specialized) across its multiple R&D locations than the distribution of the whole industry for the same locations. Interestingly, large pharma and small ventures demonstrate striking difference in this measure: -0.021 vs. 0.070, both significantly different from zero, the industry benchmark.

Finally, since citations beyond the end of the sample period are unobservable, the forward citation measure inevitably encounters data truncation problems, especially for the most recent patents (Hall *et al*., 2001). To alleviate potential biases, we apply the *year dummies* in all regressions, and conduct robustness tests using an earlier sample period. Meanwhile, because the typical number of citations a patent receives may vary across technology categories, we use the *technology dummies* to represent the 14 primary patent classes in the sample.

The regressions are based on the following equation:

$$E(local\_cite) = \beta_0 + \beta_1 \cdot cross\_region + \beta_2 \cdot big\_pharma + \beta_3 \cdot cross\_region \times big\_pharma + \beta_4 \cdot inventors + \beta_5 \cdot firm\_size + \beta_6 \cdot loc\_similarity + \beta_7 \cdot region\_size + \sigma_i + \varsigma_t \quad (2)$$

where σ_i and ς_t are vectors of technology and year dummies, respectively. We also use a cluster model to allow for the possibility that the observations are independent across firms but not within firms.

3.3.2 Model II: Quality of Local Innovations

Model I tracks the process of localized knowledge spillover. However, it does not address the actual result of such spillover: whether the spillover generated by cross-regional collaborations lead to high-quality of R&D by local inventors. In Model II, we directly test the relationship between the quality of local R&D and the cross-regional ties observed in the region.

In this model, the dependent variable is *value* – the count of all forward citations a patent receives after its grant date, excluding self-citations. Various studies have shown that forward citation count serves as a good indicator of a patent's economic and technological importance (e.g., Harhoff, *et al*., 1999; Hall *et al*., 2001, 2005), which is also positively correlated with other measures of patent value, such as consumer-surplus, patent renewal rate, and contribution to market capitalization (Singh, 2005b). This holds true even after we take strategic patenting (Hall & Ziedonis, 2001) into consideration. To reduce the positive skewness observed in the citation data, we also use *log_value* – the natural logarithm of (1 + *value*) – as an alternative measure.

There are two key independent variables: the prevalence of long-distance collaborations in the region (*connection*), and the percentage of these collaborations organized by big pharmas (*conn_bpharma*). The first variable is defined as the percentage of patents in a region that result from collaborations between local inventors and someone outside of the region. The second variable addresses the organizational context of the collaborations. It is defined as the percentage of cross-regional collaborations that are assigned to the top five percentile firms in terms of pharmaceutical patent output. As we can tell from

here, Model I pays more attention to the cross-regional ties at the firm level while Model II focuses more on the regional environment.

Similar to Model I, the control variables include the number of scientists on the team (*inventors*), the total patent output of the firm (*firm_size*) and the region (*region_size*), as well as the *year* and *technology* dummies. In addition, we want to control for two region-level variables. First is the presence of big pharma itself. Large firms are usually more resourceful in improving the local infrastructure and attracting the best talent to town. *Ceteris paribus*, their presence should enhance the overall level of local R&D. Therefore, we use the variable *big_ratio* to measure the percentage of patents in a region that are assigned to the top five percentile firms. Concerned with the correlation between *big_ratio* and *conn_bpharma*, the percentage of cross-regional collaborations organized by big pharmas, we will also run regressions without *big_ratio* and verify the robustness of the results. The second variable is *tech_overlap*, the degree of technological overlap that the region has with all other regions. Following equation (1), for each region i, $tech\_overlap_i$ is the mean of all s_{ij}'s across j. A region with a small *tech_overlap* value is probably specialized in some niche areas.

The main variables are summarized in Table 1. Note that for the highly skewed variables, such as *firm_size* and *region_size*, only the logarithmic values are reported. Table 2 presents the correlations among these variables.

Insert Tables 1 and 2 here

The regressions are based on the following equation:

$$E(value) = \beta_0 + \beta_1 \cdot connection + \beta_2 \cdot conn\_bpharma + \beta_3 \cdot inventors + \beta_4 \cdot cross\text{-}region + \beta_5 \cdot firm\_size + \beta_6 \cdot region\_size + \beta_7 \cdot big\_ratio + \beta_8 \cdot tech\_overlap + \sigma_i + \varsigma_t \quad (3)$$

where σ_i and ς_t are vectors of technology and year dummies, respectively. Similar to Model I, a cluster model is used to allow for the possibility that the observations are not independent within firms.

IV. EMPIRICAL RESULTS

In this section, we describe the main empirical results of the two alternative models, which are shown in Tables 3 and 4, respectively. The results from a series of robustness checks will be described in the next section.

4.1 Results of Model I

In the first six columns of Table 3, we examine the total number of local citations received by the focal patent, *local_cite*. To address the serious skewness in the distribution, we first take the natural logarithm of *local_cite* + 1, and use a simple OLS with technology and year dummies. The results are shown in columns (1) to (5). Column (6) is parallel (1), only that a negative binomial model is used to account for the discrete nature of citation counts.

Insert Table 3 here

The baseline results in column (1) suggest that a patent resulting from a cross-regional collaboration tends to generate more citations by local inventors. According to the marginal effect calculation, such a patent in general generate 25% more local citations than a comparable patent whose inventors are all local. However, this positive effect is seriously compromised if the cross-regional collaborations belong to one of the largest pharmaceutical companies in the world. In supportive of the theoretical arguments made in Section II, the coefficient on the interaction term *cross_region* × *big_pharma* indicates that cross-regional collaborations in big pharmas contribute less than half as much to local innovation as those formed by other organizations.

In column (2), we use the interaction term between *cross_region* and *firm_size*, the continuous measure of firm size instead of the categorical variable *big_pharma*. The same pattern holds: while the coefficient on *cross_region* remains positive and significant, the coefficient on the interaction term is negative and

significant. That is, the knowledge spillover benefit from the cross-regional collaborations may become very limited if the originating firm is of a considerably large size. Alternatively speaking, locating right next to a big pharma is not necessarily a bonus. Even though patents developed by large firms, on average, generate more local spillover, this is not the case if the large firms have intensive internal linkages across different geographic locations.

In column (3), we restrict the sample to the top five percentile regions in terms of pharmaceutical patent output in the observation year. Most of the theoretical discussions on localized knowledge spillover apply to densely-populated technology clusters where scientists frequently interact with one another. Yet, most of the regions in the sample have only a handful of patents every year, locations that do not have much to do with "technology clusters." Under the restriction, the number of distinct regions in the sample drops from 361 to 17 (7 foreign countries and 10 U.S. metropolitan areas), although they represent nearly 75% of the patent output. The results become even more significant with the restricted sample. Similar results are obtained when we restrict the sample to regions with annual pharmaceutical patent output of 100 or higher.

The factor of regional specialization enters the regression in column (4). Not surprisingly, firms that exercise more specialization than average, i.e. with a lower *loc_similarity* score, generate less spillover. Since big pharmas prove to exercise more specialization, the coefficient on the cross term *cross_region* × *big_pharma* becomes slightly smaller in magnitude, though still negative and significant. The same pattern remains even after we include regional fixed effects in column (5).

Results obtained with the negative binomial model – as shown in column (6) – are very consistent with those in the first three columns, both in the signs and magnitudes of the coefficients. The control variables also show stable results across specifications: The number of inventors on the patent, the firm size, and the level of R&D activities in the region are all positively associated with local citations received by the focal patents.

The dependent variable for columns (7) is *local_small_cite*, which counts the number of citations that the focal patent receives from local small ventures – those with less than five patents during the observation year. The only difference from the previous columns is that the coefficient on *firm_size* changes from significantly positive to insignificantly negative, suggesting the possibility that firms are more likely to cite technologies generated from similar organizations.

Because cross-regional collaborations involve more than one location, the resulting patents would also call multiple regions "local." Although for any particular "neighbor," what matters is the knowledge spillover at one specific location, it is still interesting to see whether the aggregated numbers will present a different picture. In column (8), we sort out all the regions that each patent has inventors in, and sum up the citations from all these locations to measure the localized spillover. As expected, the coefficient on *cross_region* gets much higher after the aggregation, but the coefficient on the interaction term remains negative and significant, indicating the internalization effect of cross-regional collaborations within big pharmas.

4.2 Results of Model II

In Model II, the overall quality of local R&D is associated with the cross-regional connections found in the local community. Similar to Model I, we apply two parallel regression methods: OLS on the logged dependent variable, and negative binomial on the direct citation counts.

Insert Table 4 here

Due to the relatively high correlation between *big_ratio*, the percentage of local patents generated by big pharmas, and *conn_bpharma*, the percentage of cross-regional ties attributable to big pharmas, we leave out *big_ratio* in the regression in column (1) before putting it into the regression in column (2). The technological overlap variable *tech_overlap* is added in column (3), and both *big_ratio* and *tech_overlap* are

added in column (4). Moreover, different firms are affected by local innovation environments to various degrees. For example, without the functional support of established internal organizations, small ventures are most likely affected by the activities of neighboring entities. Hence, we run the regressions on a subset of patents that are developed by firms with fewer than five patents in the observation year, and report the results in column (5). Columns (6) and (7) are simply replications of columns (1) and (4), respectively, with negative binomial regressions.

Consistently, extensive cross-regional ties contribute positively to the quality of local innovation, with the strongest effect occurring to the patents of small ventures. For an average local patent, increasing the value of *connection* by one standard deviation will increase the expected forward citation counts by around 50%. Even after controlling for a series of regional characteristics, well-connected locations tend to produce patents with a larger impact.

However, if the cross-regional ties are mostly formed by large pharmas, the benefits may be significantly reduced, if not completely wiped out. In fact, for a region with the average *connection* level at 35% and *conn_bpharma* at 30%, having additional cross-regional collaborations formed by large pharmas may only make things worse. This observation is consistent with the influence of firm organization: cross-regional ties within a large, established firm may also serve the purpose of building a closely-knit internal R&D network, making knowledge less decipherable by the neighboring inventors.

The coefficients on the control variables are consistent with theoretical conjectures. At the patent level, the positive coefficients on *inventors* indicate that having more inventors on the team is associated with higher value of the innovation. It also helps to have inventors from different geographic areas, as evidenced by the strong positive coefficients on *cross-region*. At the firm level, firm size seems to have a negative effect on patent value except among the very small firms. This may be due to large firms' higher propensity of patenting as suggested by some early studies (Kortum & Lerner, 1998; Hall & Ziedonis, 2001).

At the location level, patents developed in larger technology clusters receive more forward citations. This is in line with the localized knowledge spillover argument: because learning is more likely to happen at the local level, and because there are more potential learners in technology clusters, patents generated in clusters tend to receive more citations in the future. The strong presence of big pharmas in a region, i.e., a large *big_ratio*, is positively associated with the value of local innovations, indicating the spillover effect from the strongest players in the industry. Acs & Audretsch (1988) reach similar findings in their multi-industry studies. Finally, the coefficient on *tech_overlap* is positive and significant, suggesting that highly specialized regions may be less likely to produce high-impact innovation.

V. ROBUSTNESS TESTS AND DISCUSSIONS

In the above section, we find strong empirical evidence that cross-regional collaborations are generally associated with more spillover to the local community and more valuable local innovations. However, cross-regional collaborations formed by the largest companies show significantly lower spillover effects. In this section, a series of robustness tests are conducted to make sure that the previous findings are not dependent on the specific setups.

5.1 Robustness tests for Model I

The robustness tests on local knowledge spillover are illustrated in Table 5.

Insert Table 5 here

First, administrative boundaries are only approximate descriptions of R&D locations. In the first four columns of Table 5, we use two alternative definitions of "region": the state boundaries, and the economic areas following the Bureau of Economic Analysis (BEA). Country boundaries are still used for the rest of the world. These two alternative definitions, respectively, generate 149 and 268 unique geographic

regions for the sample. The regress results, with both the 0-1 *big_pharma* measure and the continuous *firm_size* measure, remain strong and significant.

Column (5) addresses the concern over the excessive number of zero local citations. In fact, over half of the patents in the sample had never received any citations from the local community by the end of the sample period, which may be due to the short observation window or the patents' low intrinsic value rather than firms' strategic knowledge internalization. Hence, a zero-inflated negative binomial model is applied to allow for the alternative mechanism for zero local citation, with the total number of forward citations as the exposure variable. The result does confirm the possibility that the observation window is a significant predictor of the excessive zeroes in the dependent variable, but we find no significant changes to the coefficients of the key variables.

One caveat in this study is that firm organizations are not exogenous; instead, firms strategically organize their R&D activities in response to the external environments. For example, locations with inadequate supply of human capital and weak technology base will find it more difficult to attract industry leaders to set up R&D centers there. Once they do, the firms' local subsidiaries are more likely to maintain close connections with headquarters for technical support, connections that would certainly include cross-regional collaborations. Thus, the low local citation counts may simply reflect the fact that there is not much serious R&D going on. In column (6), we allow for endogeneity of the focal variable *cross-region* – as well as the cross term – by applying a three-stage least squares regression, where the decision to engage in cross-regional collaboration is dependent on the location characteristics and the amount of R&D the firm has locally. The results support the argument that non-cluster areas are more likely to see cross-regional collaborations, and the results with the focal variables get even stronger.

Next, we test the appropriate definition of big pharma. Since the analytical focus here is on the role of firm organization, we want to pay particular attention to the way firms are characterized. In the baseline analysis, firm size is measured by the total number of patents filed in the observation year. Although the

overall size is important, it may not reflect the R&D activities a firm carries out at a specific location. In column (7), we redefine big pharmas as those (i) belonging to the top five percentile in terms of global patent output, and (ii) having at least 20% of their patents developed locally. The results remain strong. It is worth mentioning that, while the coefficients on *connection* and the cross term remain unchanged, the coefficient on *big_pharma* becomes positive and significant, indicating that a local giant with substantial R&D in the community contribute more to local innovation than a "listening post" set up by a large multi-location firm.

Finally, the count of forward citations received by the focal patent is highly dependent on the time horizon in which the citations are observed. Hall *et al*. (2001) shows that it took ten years for the 1975 patents to receive 50% of their forward citations. Even with the year fixed effects, the dependent variable can still be too noisy a proxy for the value of the most recent patents. In column (8), we conduct the analysis on patents granted before 1995, leaving us at least ten years of observation window. All the key results remain with this much smaller sample.

5.2 Robustness tests for Model II

The robustness tests on the value of local patents are illustrated in Table 6. Here, OLS on the logged dependent variable is used except in column (5).

Insert Table 6 here

Similar to Table 5, the first four columns in Table 6 test the alternative definitions of regions using state and economic area boundaries, respectively. Regressions both with and without the *big_ratio* variable are presented. The results are highly consistent with the findings with metropolitan area boundaries. Column (5) addresses the excessive observations of zeros in the dependent variable. Nearly one third of the patents in the sample never received any citations other than from the innovating firms themselves, and

again, the short observation window may be to blame. After allowing the alternative mechanism for the zero observations, we find the same strong results as in Table 4.

Endogeneity issue is again raised in column (6), where two key variables are treated as endogenous. First, as in Model I, the decision to form cross-regional collaborations may be dependent on the availability of local resources. Second, the percentage of cross-regional ties attributable to big pharmas may be determined by the sheer dominance of big pharmas in local R&D. Both these conjectures are supported by the simultaneous equations, but the coefficients in the main equation are still significant with the expected signs.

Lastly, to address the truncation problem, we break the sample into two periods: the first from 1975 to 1994, and the second from 1995 to 2001. The main results hold in both sample periods, confirming the robustness of the findings. The negative moderating effect of big pharma seems to be more significant in the second period, implying that the fast development of information technologies in the past decade may have helped with the global firms' internalization efforts more than the promotion of knowledge spillover. Of course, we should exercise caution in interpreting such results, given the potential truncation problem for the second sample period.

5.3 Discussions on the empirical results

Admittedly, there are still many limitations to the empirical findings. First, knowledge spillover is not restricted to science and technology, not to mention patented technologies. Large firms located in certain areas may help spread information on markets, regulations, and managerial practice, which directly or indirectly affects the neighboring firms' innovation activities. These effects cannot be fully captured in this study. More information is needed for a comprehensive understanding of local interactions.

In addition, to truly capture the effect of cross-regional collaborations on knowledge spillover, we need to take inventor mobility (Almeida & Kogut, 1999) into consideration, both within and across firm

boundaries. For instance, due to internal mobility, R&D carried out at one location may be recorded as cross-regional collaboration when the patent is filed several years later, and vice versa. Moreover, a former employee of an incumbent firm may start up a new business (Agarwal *et al.*, 2004) outside of the region. In such circumstances, patents resulting from a previous collaboration may appear as cross-regional collaborations, although they hardly reflect the incumbent firm's strategic internal organization.

Most importantly, the above analysis does not indicate any causal relationships among the key variables. To certain extent, the first-hand observations during interviews complement the empirical analyses in showing that organizations play an important role in influencing knowledge flow and knowledge spillover, beyond what can be explained by the ties among individuals. Nevertheless, many more dimensions – such as firms' strategic location decisions and acquisition activities – need to be sorted out before we can talk about causality with any certainty.

VI. CONCLUSION

There has been a large literature on the role of interpersonal networks on knowledge transfers. It is far less understood how this role is actively shaped by the organizational context around them. By taking a closer look at a particular type of interpersonal ties – cross-regional collaborations among pharmaceutical researchers – we argue that firm organization has a significant impact on the relationship between interpersonal networks and knowledge flow. While collaborations with the outside world are generally beneficial to local R&D, the benefit can be significantly reduced if these connections are formed within large, established organizations.

Going back to the example discussed at the very beginning, the findings in this study suggest that the Shanghaiese do have reason to celebrate the entry Roche's new R&D center. Industry leaders in a region usually generate more knowledge spillover and contribute to the value of local R&D. However, knowledge spillover from Roche can be very limited if the strong linkages among its multiple R&D

centers also promote knowledge internalization, hence raising the learning barrier faced by the outsiders. In particular, if most of a region's external interactions are dependent on the networks within large multinational firms, other firms may expect very little from the localized spillover.

This study also contributes to the organization literature by suggesting that organizational structures not only affect knowledge management directly (Argyres & Silverman, 2004), but also do so indirectly by influencing the functioning of interpersonal ties. Moreover, the implications of such interactions extend beyond the organizational boundaries. Although the emphasis on firm organization has resulted in the simplification on other important dimensions, such as the strength of the collaborative ties (Hansen, 1999) and the nature of the knowledge being transferred (Zander & Kogut, 1995; Hansen *et al*., 2005), this study is complementary to many of the findings in this literature.

Finally, a deeper understanding of this phenomenon will have important implications to policy makers who are eager to attract investments and nurture local technology clusters. With firms' geographically dispersed R&D activities, the same scale of local R&D may generate very different knowledge spillover to the neighbors, depending on how knowledge is organized internally. For the local community, being part of a multinational firm's "global network" is not always a blessing.

REFERENCES

Acs, Z.J & Audretsch, D.B. 1988. Innovation in large and small firms: An empirical analysis. ***American Economic Review***, 78(4): 678–690

Agarwal, R., Echambadi, R., Franco, A. & Sarkar, M.B. 2004. Knowledge transfer through inheritance: Spin-out generation, development and performance. ***Academy of Management Journal***, 47(4): 501–522.

Alcacer, J. & Gittelman, M. 2005. How do I know what you know? Patent examiners and the generation of patent citations. ***Review of Economics and Statistics***, forthcoming.

Alcacer, J., Islam, M. & Zhao, M. 2005. Global teams and local knowledge flows: Are large firms contributing to innovation more than we thought? Working paper.

Almeida, P. & Kogut, B. 1999. Localization of knowledge and the mobility of engineers in regional networks. ***Management Science,*** 45(7): 905–917.

Anand, B.N. & Galetovic, A. 2004. How market smarts can protect property rights. ***Harvard Business Review,*** 82 (12):72–79.

Argyres, N.S. & Silverman, B.S. 2004. R&D, organization structure, and the development of corporate technological knowledge. ***Strategic Management Journal***, 25(8-9): 929–958.

Arora, A. & Gambardella, A. 1994. The changing technology of technological change: general and abstract knowledge and the division of innovative labor. ***Research Policy***, 23(5): 523–532.

Audretsch, D.B. & Feldman, M.P. 1996. R&D spillovers and the geography of innovation and production. ***American Economic Review,*** 86(3): 630–640.

Bartlett, C.A. & Ghoshal, S. 1990. The multinational corporation as an interorganizational network. ***Academy of Management Review***, 15(4): 603–625.

Buckley, P.J. & Casson, M.C. 1976. ***The future of the multinational enterprise***. London: Homes & Meier.

Cantwell, J.A. & Janne, O. 1999. Technological lobalization and innovative centres: The role of corporate technological leadership and locational hierarchy. ***Research Policy***. 28: 119–144.

Chandler, A.D., Jr. 2005. ***Shaping the industrial century: The remarkable story of the evolution of the modern chemical and pharmaceutical industries.*** Cambridge, MA: Harvard University Press.

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

Cohen, M.D. 2006. What's different is routine. ***Industrial and Corporate Change***, 15(2): 387–390.

Cohen, W.M. & Levinthal, D. 1990. Absorptive capacity: A new perspective on learning and innovation. ***Administrative Science Quarterly***, 24: 128–152

Cohen, W.M., Nelson, R.R. & Walsh, J.P. 2000. Protecting their intellectual assets: Appropriability conditions and why U.S. manufacturing firms patent (or not). ***NBER Working Paper***, No. 7552.

Dierickx, I. & Cool, K. 1989. Asset stock accumulation and sustainability of competitive advantage. ***Management Science***, 35 (12): 1504−1511.

Feinberg, S.E. & Gupta, A.K. 2004. Knowledge spillovers and the assignment of R&D responsibilities to foreign subsidiaries. ***Strategic Management Journal***, 25: 823–845.

Fleming, L. 2001. Recombinant uncertainty in technological search. ***Management Science,*** 47(1): 117–132.

Fleming, L., King, C. III & Juda, A. 2004. ***Small Worlds and Innovation.*** Working paper. Harvard Business School**.**

Frost, T.S. & Zhou, C. 2005. R&D co-practice and 'reverse' knowledge integration in multinational firms. ***Journal of International Business Studies***, 36(6): 676–687.

Furman, J.L., Kyle, M.K., Cockburn, I. & Henderson, R. 2004. Public and private spillovers, location, and the productivity of pharmaceutical research. Forthcoming in ***Annales d'Economie et Statistique***.

Hall, B.H., Jaffe, A.B. & Trajtenberg, M. 2001. The NBER patent citation data file: lessons, insights and methodological tools. ***NBER Working Paper***, No. 8498. Also in: Jaffe and Trajtenberg (Eds). 2002. ***Patents, Citations, and Innovations***. Cambridge: MIT Press.

Hall, B.H., Jaffe, A.B. & Trajtenberg, M. 2005. Market value and patent citations. ***Rand Journal of Economics***, 36: 16−38.

Hall, B.H. & Ziedonis, R.H. 2001. The patent paradox revisited: An empirical study of patenting in the U.S. ***The Rand*** Journal of Economics, 32(1): 101–128.

Hansen M.T. 1999. The search-transfer problem: the role of weak ties in sharing knowledge across organization subunits. ***Administrative Science Quarterly***, 44 (1): 82–111.

Hansen, M.T., Mors, M.L. & Lovas, B. 2005. Knowledge Sharing in Organizations: Multiple networks, multiple Phases. ***Academy of Management Journal***, 48(5): 764−793

Harhoff, D., Narin, F., Scherer, F.M. & Vopel, K. 1999. Citation frequency and the value of patented inventions. ***Review of Economics and Statistics***, 81(3): 511–515.

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

Henderson, R.M. & Cockburn, I. 1996. Scale, scope, and spillovers: The determinants of research productivity in drug discovery. ***The RAND Journal of Economics,*** 27: 32–59.

Hausman, J., Hall, B.H. & Griliches, Z. 1984. Economic models for count data with an application to the patent-R&D relationship. ***Econometrica***, 52(4): 909–938.

Jaffe, A.B. 1986. Technological opportunity and spillovers of R&D: Evidence from firms' patents, profits, and market value. ***American Economic Review***, 76(5): 984–1001.

Jaffe, A. B., Trajtenberg, M. & Henderson, R. 1993. Geographic localization of knowledge spillovers as evidenced by patent citations. ***Quarterly Journal of Economics,*** 434: 578–98.

Jaffe, A. B., Trajtenberg, M. & Fogarty, M.S. 2000. The meaning of patent citations: Report on the NBER/Case-Western Reserve survey of patentees. ***NBER Working Paper***, No. 7631.

Jones, B.F. 2005. The burden of knowledge and the 'death of the Renaissance man': Is innovation getting harder? ***NBER Working Paper*** No. 11360

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

Kortum, S. & Lerner, J. 1998. Stronger protection or technological revolution: What is behind the recent surge in patenting? ***Carnegie-Rochester Conference Series on Public Policy***, 48: 247–304.

Lahiri, N. 2003. Inter-firm technology diffusion: Do intra-organizational linkages matter? ***Doctoral dissertation***, Ross School of Business, University of Michigan.

Leamer, E.E. & Storper, M. 2001. The economic geography of the Internet age. ***NBER Working paper***, No. 8450.

Levin, R.C. 1988. Appropriability, R&D spending, and technological performance. ***American Economic Review***, 78 (2): 424–428.

Levin, R.C., Klevorick, A.K., Nelson, R.R., & Winter, S.G. 1987. Appropriating the returns from industrial R&D. ***Brookings Papers on Economic Activity***, 3: 783–820.

MacGarvie, M. & Furman, J. L. 2005. Early academic science and the birth of industrial research laboratories in the US pharmaceutical industry. ***NBER Working Paper***, No. 11470.

March, J. 1991. Exploration and exploitation in organizational learning. ***Organization Science***, 2: 71–87.

McEvily, B. & Marcus, A. 2005. Embedded ties and the acquisition of competitive capabilities. ***Strategic Management Journal***, 26(11): 1033–1055.

Nelson, R. & Winter, S. 1982. ***An Evolutionary Theory of Economic Change***. Harvard University Press: Cambridge, MA.

Penner-Hahn, J. & Shaver, J.M. 2005. Does international research and development affect patent output? An analysis of Japanese pharmaceutical firms. ***Strategic Management Journal***, 26(2): 121–140.

Podolny, J. & Stuart, T.E. 1995. A role-based ecology of technological change. ***American Journal Sociology***, 100(5): 1224–1260.

Rajan R. & Zingales, L. 2001. The firm as a dedicated hierarchy: A theory of the origins and growth of firms. ***Quarterly Journal of Economics***, 116 (3): 805–851.

Reagans, R. & Zuckerman, E.W. 2001. Networks, diversity and productivity: The social capital of corporate R&D teams. ***Organization Science***, 12(4): 502–517.

Rivkin, J.W. 2000. Imitation of complex strategies. ***Management Science***, 46(6): 824–844.

Schumpeter, J.A. 1939. ***Business Cycles***. New York: McGraw-Hill.

Scott, W.R. & Davis, G.F. 2006. ***Organizations and Organizing: Rational, Natural and Open Systems Perspectives***, 6th edition. New Jersey: Prentice Hall.

Simon, H.A. 1997. ***Administrative Behavior***, 4th edition. Free Press.

Singh, J. 2005a. Collaborative networks as determinants of knowledge diffusion patterns. ***Management Science,*** 51(5): 756–770.

Singh, J. 2005b. Distributed R&D, cross-regional ties and quality of innovative output. Working paper.

Sonn, J.W. & Storper, M. 2004. The increasing importance of geographical proximity in technological innovation: An analysis of US patent citations, 1975–1997. Working paper.

Szulanski, G. 1996. Exploring internal stickiness: Impediments to the transfer of best practice within the firm. ***Strategic Management Journal***, 17(Winter): 27–43.

Teece, D.J. 1986. Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy. ***Research Policy,*** 15(6): 285–305.

Uzzi, B. & Spiro, J. 2005. Collaboration and creativity: The small world problem. ***American Journal of Sociology***, 111(2): 447–504.

Verspagen, B. & Schoenmakers, W. 2004. The spatial dimension of patenting by multinational firms in Europe, ***Journal*** of Economic Geography, 4(1): 23–42.

Zander, U. & Kogut, B. 1995. Knowledge and the speed of the transfer and imitation of organizational capabilities', ***Organization Science***, 6(1):76–92.

Zhao, M. 2006. Conducting R&D in countries with weak intellectual property rights protection. ***Management Science***, 52(8): 1185–1199.

Zucker, L.G. & Darby, M.R. 2005. An evolutionary approach to institutions and social construction: Process and structure. In Smith & Hitt (Eds). ***The Oxford Handbook of Management Theory: The Process of Theory Development***. Oxford, UK: Oxford University Press.

Figure 1. Firms are spreading out geographically

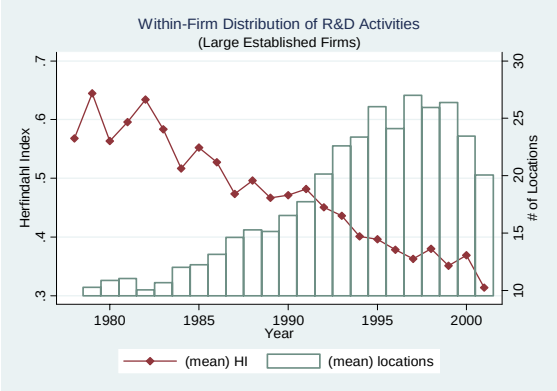
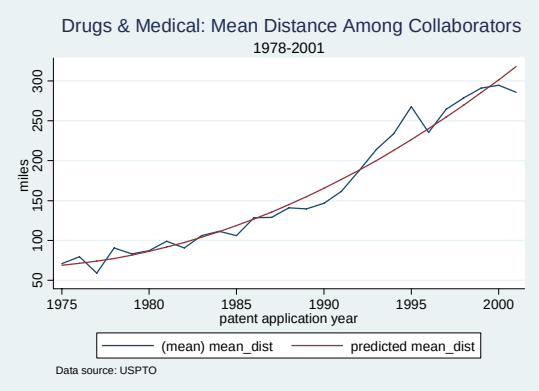


Figure 2. More cross-regional collaborations



- Here the criteria for large established firms are (1) top% percentile in terms of patent output, and (2) continuously in business for at least 25 years.
- Locations are defined as metropolitan areas in the U.S. and countries outside of the U.S. The same pattern remains with other definitions of location.

Figure 3. Internal information sharing across distance

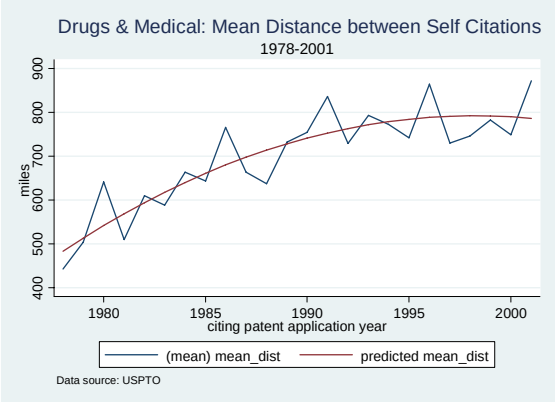
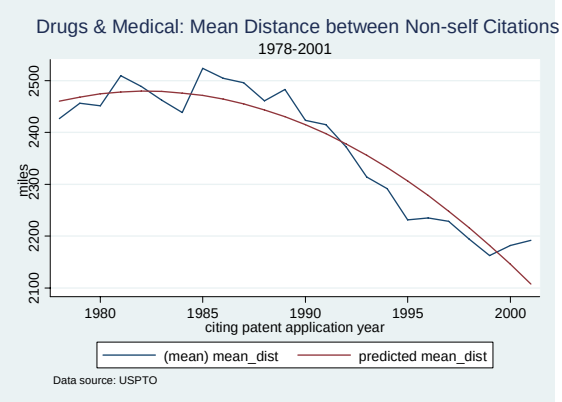


Figure 4. Knowledge spillover across firms



- Citation distance is defined as the mean of the pairwise distances between all inventors on the citing patent and all inventors on the cited patent.

Table 1. Summary of Variables

Variable		Explanation	Mean	Std. Dev.	Min.	Max.
Patent level	*value*	Number of citations a patent receives, excluding self citations	6.27	14.84	0	1,075
	local_cite	Number of citations a patent receives from the same region, excluding self citations	1.47	5.11	0	305
	local_small_cite	Number of citations from local small companies, excluding self citations	0.46	1.79	0	109
	inventors	Number of inventors on the patent	2.55	1.73	1	32
	cross_region	1 if the patent is developed through cross-regional collaboration, 0 otherwise	0.17	0.38	0	1
Firm level	*firm_size*	Natural logarithm of 1 + the total number of patents filed by the firm	0.80	0.30	0.69	4.67
	loc_similarity	Similarity of technology classes across the firm's multiple locations	0.03	0.22	-0.93	0.99
Regional level	*region_size*	Natural logarithm of 1 + the total number of patents developed in the region	1.41	1.06	0.69	6.33
	connection	Percentage of all patents with inventors from multiple locations	0.35	0.42	0	1
	conn_bpharma	Percentage of cross-regional collaborations that are in big pharmas	0.29	0.41	0	1
	big_ratio	Percentage of patents granted to large firms	0.16	0.32	0	1
	tech_overlop	The degree of technological overlap with other regions	0.54	0.13	0	0.65
tech fixed effect		Dummy variables for the 14 primary patent classes				
year fixed effect		Dummy variables for the 30 grant years				

Table 2. Correlation Matrix

		1	2	3	4	5	6	7	8	9	10	11	12
1	*value*	1.00											
2	*local_cite*	0.58	1.00										
3	*local_small_cite*	0.55	0.80	1.00									
4	*inventors*	-0.04	0.02	0.03	1.00								
5	*cross_region*	-0.01	0.02	0.06	0.35	1.00							
6	*firm_size*	0.00	-0.01	-0.06	0.08	-0.12	1.00						
7	*loc_similarity*	0.05	0.03	0.01	-0.02	0.10	-0.20	1.00					
8	*region_size*	-0.10	0.00	0.02	0.08	-0.13	0.18	-0.19	1.00				
9	*connection*	-0.07	-0.05	0.00	0.05	0.34	-0.03	0.08	-0.31	1.00			
10	*conn_bpharma*	-0.08	0.00	-0.03	0.08	-0.05	0.31	-0.08	0.26	-0.05	1.00		
11	*big_ratio*	-0.06	0.02	-0.01	0.02	-0.08	0.37	-0.04	0.28	-0.03	0.78	1.00	
12	*tech_overlap*	-0.08	-0.01	0.00	0.07	-0.05	0.02	-0.12	0.16	-0.07	0.06	0.07	1.00

Table 3. Regressions on Local Knowledge Spillover

Dependent Variable: *local_ cite* and *local_small_cite*

	Citations from all local inventors						**Citations by local small firms**	**Citations from multi-regions**
	OLS on log (1+ DV)					Negative Binomial		
			cluster only		location FE			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
cross_region	0.155 ** (0.011)	0.201 ** (0.015)	0.156 ** (0.013)	0.147 ** (0.011)	0.113 ** (0.005)	0.470 ** (0.044)	0.109 ** (0.008)	0.389 ** (0.014)
big_pharma	- 0.002 (0.020)		0.006 (0.025)	- 0.007 (0.020)	- 0.004 (0.006)	- 0.045 (0.071)	- 0.016 (0.013)	- 0.006 (0.022)
cross_region × *big_pharma*	- 0.080 ** (0.017)		- 0.098 ** (0.021)	- 0.073 ** (0.017)	- 0.057 ** (0.006)	- 0.243 ** (0.063)	- 0.058 ** (0.011)	- 0.085 ** (0.022)
inventors	0.024 ** (0.003)	0.024 ** (0.003)	0.023 ** (0.001)	0.024 ** (0.003)	0.028 ** (0.001)	0.065 ** (0.010)	0.011 ** (0.002)	0.037 ** (0.004)
firm_size	0.040 ** (0.008)	0.042 ** (0.007)	0.036 ** (0.010)	0.042 ** (0.008)	0.028 ** (0.002)	0.139 ** (0.026)	- 0.005 (0.005)	0.042 ** (0.009)
cross_region × *firm_size*		- 0.034 ** (0.006)						
loc_similarity				0.109 ** (0.040)	0.036 ** (0.002)			
region_size	0.042 ** (0.004)	0.042 ** (0.004)	0.023 (0.016)	0.043 ** (0.004)	- 0.059 ** (0.005)	0.133 ** (0.013)	0.028 ** (0.002)	0.044 ** (0.004)
constant	0.147 (0.092)	0.141 (0.093)	0.226 † (0.132)	0.132 (0.094)	0.668 ** (0.136)	- 1.059 ** (0.373)	0.050 (0.035)	- 0.141 (0.094)
technology dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	247,127	247,127	157,032	247,127	238,110	247,127	247,127	247,127
F or χ^2	59.89	56.12	47.81	58.92	746.17	4,174.60	37.20	66.28
Prob >F or χ^2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Numbers in parentheses ( ) are robust standard errors allowing for within-firm correlation except (5).

† $p < 0.10$
* $p < 0.05$
** $p < 0.01$

Table 4. Regressions on the Value of Local Patents

Dependent Variable: *value*

	OLS on log (1+ DV)					Negative Binomial	
					small firm patents		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
connection	0.506 ** (0.042)	0.476 ** (0.043)	0.535 ** (0.046)	0.504 ** (0.048)	0.585 ** (0.037)	0.885 ** (0.083)	0.875 ** (0.088)
conn_bpharma	- 0.288 ** (0.039)	- 0.463 ** (0.043)	- 0.291 ** (0.040)	- 0.460 ** (0.043)	- 0.456 ** (0.036)	- 0.361 ** (0.052)	- 0.639 ** (0.062)
inventors	0.007 * (0.003)	0.008 * (0.003)	0.007 * (0.003)	0.007 * (0.003)	- 0.006 † (0.003)	0.024 ** (0.006)	0.025 ** (0.006)
cross_region	0.081 ** (0.011)	0.082 ** (0.011)	0.083 ** (0.011)	0.084 ** (0.011)	0.108 ** (0.012)	0.133 ** (0.023)	0.139 ** (0.023)
firm_size	- 0.026 ** (0.007)	- 0.030 ** (0.007)	- 0.026 ** (0.007)	- 0.030 ** (0.007)	0.064 ** (0.014)	- 0.033 ** (0.013)	- 0.041 ** (0.013)
region_size	0.018 ** (0.005)	0.015 * (0.006)	0.019 ** (0.004)	0.015 ** (0.006)	0.022 ** (0.004)	0.042 ** (0.008)	0.034 ** (0.009)
big_ratio		0.239 ** (0.057)		0.230 ** (0.058)	0.317 ** (0.043)		0.395 ** (0.083)
tech_overlap			0.140 ** (0.044)	0.129 ** (0.044)	0.176 ** (0.032)		0.242 ** (0.075)
constant	1.675 ** (0.140)	1.680 ** (0.143)	1.606 ** (0.144)	1.616 ** (0.146)	1.182 ** (0.291)	1.730 ** (0.211)	1.637 ** (0.218)
technology dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	244,919	244,919	244,919	244,919	91,697	244,919	244,919
F or χ^2	315.26	313.25	308.61	306.97	697.71	16,551.32	16,942.25
Prob >F or χ^2	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Numbers in parentheses ( ) are robust standard errors allowing for within-firm correlation.

† $p < 0.10$
* $p < 0.05$
** $p < 0.01$

Table 5. Robustness Tests on Local Knowledge Spillover

Dependent Variable: *local_ cite*

	Alternative definition of regions				**ZINB**	**Endogenous variable**	**Redefine big pharma**	**Granted < 1995**
	State (1)	State (2)	EA (3)	EA (4)	(5)	(6)	(7)	(8)
cross_region	0.183 ** (0.011)	0.531 ** (0.040)	0.175 ** (0.011)	0.315 ** (0.028)	0.221** (0.026)	1.988 ** (0.102)	0. 153 ** (0.010)	0.242 ** (0.018)
big_pharma	0.002 (0.021)		- 0.043 (0.021)		0.004 (0.054)	0.512 ** (0.034)	0.085 ** (0.017)	0.003 (0.034)
cross_region × *big_pharma*	- 0.096 ** (0.017)		- 0.089 ** (0.017)		- 0.210 ** (0.041)	- 2.283 ** (0.106)	- 0.087 ** (0.019)	- 0.085 * (0.034)
inventors	0.021 ** (0.003)	0.020 ** (0.003)	0.021 ** (0.003)	0.020 ** (0.003)	0.058 ** (0.006)	- 0.014 ** (0.005)	0.024 ** 0.002	0.022 ** (0.004)
firm_size	0.030 ** (0.008)	0.020 ** (0.007)	0.038 ** (0.008)	0.027 ** (0.006)	0.187 ** (0.018)	- 0.065 ** (0.002)	0.014 ** (0.005)	0.064 ** (0.018)
firm_size × *big_pharma*		- 0.065 ** (0.007)		- 0.032 ** (0.005)				
region_size	0.085 ** (0.005)	0.108 ** (0.006)	0.052 ** (0.004)	0.064 ** (0.005)	0.132 ** (0.011)	0.081 ** (0.003)	0.041 ** (0.004)	0.060 ** (0.007)
constant	0.017 (0.095)	- 0.078 (0.098)	0.141 ** (0.097)	0.102 (0.099)	- 2.417 ** (0.619)	- 0.327 * (0.135)	0.178 † (0.095)	- 0.066 (0.105)
technology dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Exposure					*total_citation*			
Inflate								
year					- 0.171 ** (0.019)			
total citation					0.002 * (0.001)			
Endogenous *cross_region*=								
f(*cluster,*						- 0.098 ** (0.002)		
firm local patents)						- 0.002 ** (0.001)		
Observations	245,058	245,058	243,167	243,167	159,951	247,127	247,127	91,003
F or χ^2	66.43	60.54	64.19	60.01	2,041.07	37,761.81	64.57	20.00
Prob >*F* or χ^2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Numbers in parentheses ( ) are robust standard errors allowing for within-firm correlation except (6).

† $p < 0.10$ * $p < 0.05$ ** $p < 0.01$

Table 6. Robustness Tests on the Value of Local Patents

Dependent Variable: *value*

	Alternative definition of regions				**ZINB**	**Endogenous variable**	**Granted < 1995**	**Granted >= 1995**
	State (1)	State (2)	EA (3)	EA (4)	(5)	(6)	(7)	(8)
connection	0.557 ** (0.048)	0.565 ** (0.051)	0.601 ** (0.049)	0.574 ** (0.052)	0.843** (0.083)	0.118 ** (0.028)	0.586 ** (0.071)	0.343 ** (0.041)
conn_bpharma	- 0.474 ** (0.042)	- 0.412 ** (0.050)	- 0.334 ** (0.049)	- 0.432 ** (0.049)	- 0.613 ** (0.065)	- 0.101 ** (0.022)	- 0.302 ** (0.049)	- 0.469 ** (0.052)
big_ratio		0.005 (0.056)		0.189 ** (0.065)	0.398 ** (0.082)		0.060 (0.060)	0.534 ** (0.058)
inventors	0.007 ** (0.003)	0.007 * (0.003)	0.007 * (0.003)	0.007 * (0.003)	0.025 ** (0.006)	0.007 ** (0.001)	- 0.002 (0.005)	0.013 ** (0.003)
cross_region	0.081 ** (0.012)	0.081 ** (0.012)	0.082 ** (0.012)	0.082 ** (0.012)	0.134 ** (0.023)	0.098 ** (0.005)	0.110 ** (0.020)	0.061 ** (0.011)
firm_size	- 0.030 ** (0.007)	- 0.030 ** (0.007)	- 0.027 ** (0.007)	- 0.030 ** (0.007)	- 0.042 ** (0.013)	- 0.029 ** (0.002)	- 0.051 ** (0.010)	- 0.017 ** (0.006)
region_size	0.046 ** (0.006)	0.047 ** (0.007)	0.026 ** (0.006)	0.022 ** (0.007)	0.034 ** (0.009)		- 0.004 (0.008)	0.022 ** (0.005)
constant	1.622 ** (0.145)	1.547 ** (0.148)	1.649 ** (0.140)	0.628 ** (0.143)	1.699 ** (0.217)	1.749 ** (0.167)	1.890 ** (0.154)	1.858 ** (0.054)
technology dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Inflate								
year					1.590 ** (0.178)			
Endogenous								
cross_region = f (region_size)						- 0.029 ** (0.001)		
conn_bpharma = f (big_ratio)						0.769 ** (0.001)		
Observations	242,744	242,744	240,040	240,040	244,919	247,127	89,159	155,760
F or χ^2	327.28	313.79	321.45	313.30	12,079.03	181,569.99	145.19	248.96
Prob >F or χ^2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Numbers in parentheses ( ) are robust standard errors allowing for within-firm correlation except (6).

† $p < 0.10$ * $p < 0.05$ ** $p < 0.01$

Appendix 1. Top 30 Pharmaceutical Companies in 2004

Rank	Company	Country	Healthcare Sales (US$ MN)	Healthcare R&D (US$ MN)	Employees
1	Pfizer	USA	52,516	7,684	115,000
2	Johnson & Johnson	USA	47,348	5,203	109,900
3	GlaxoSmithKline	UK	37,318	5,204	100,619
4	Sanofi-Aventis	France	31,615	4,927	96,439
5	Novartis	Switzerland	28,247	4,207	81,392
6	Hoffmann-La Roche	Switzerland	25,163	4,098	64,703
7	Merck & Co.	USA	22,939	4,010	62,600
8	AstraZeneca	UK	21,427	3,803	64,200
9	Abbott Laboratories	USA	19,680	1,697	50,600
10	Bristol-Myers Squibb	USA	19,380	2,500	43,000
11	Wyeth	USA	17,358	2,461	51,401
12	Eli Lilly and Company	USA	13,858	2,591	44,500
13	Bayer	Germany	10,554	1,299	113,060
14	Amgen	USA	10,550	2,028	14,400
15	Boehringer Ingelheim	Germany	10,146	1,532	35,529
16	Baxter International	USA	9,509	517	48,000
17	Takeda Pharmaceutical Co.	Japan	9,330	1,285	14,510
18	Schering-Plough	USA	8,272	1,607	30,500
19	Astellas Pharma	Japan	7,904	1,213	15,500
20	Procter & Gamble	USA	7,786		110,000
21	Schering	Germany	5,103	1,143	26,131
22	Merck KGaA	Germany	5,018	611	28,877
23	Eisai Co.	Japan	4,857	744	8,295
24	Novo Nordisk	Denmark	4,847	727	20,285
25	Teva Pharmaceutical Industries	Israel	4,799	338	13,813
26	Genentech	USA	4,621	948	7,646
27	Sankyo Co.	Japan	4,329	822	11,444
28	Akzo Nobel	The Netherlands	4,037	644	61,400
29	Alcon	Switzerland	3,914	390	12,200
30	Forest Laboratories	USA	3,160	294	5,136
31	Daiichi Pharmaceutical Co.	Japan	2,964	546	7,333
32	Chugai Pharmaceutical Co.	Japan	2,833	463	5,327
33	Taisho Pharmaceutical	Japan	2,655	221	5,378
34	Altana	Germany	2,623	506	10,783
35	Serono	Switzerland	2,458	595	4,902
36	Bausch & Lomb	USA	2,232	163	12,400
37	Mitsubishi Pharma	Japan	2,226	480	5,917
38	Biogen Idec	USA	2,210	684	4,266
39	Genzyme	USA	2,201	392	7,100
40	Solvay	Belgium	2,170	366	29,300
41	UCB	Belgium	2,088	404	11,403
42	Allergan	USA	2,046	346	5,030
43	Kyowa Hakko Kogyo Co.	Japan	2,035	230	5,960
44	Shionogi & Co.	Japan	1,862	279	5,522
45	Ivax	USA	1,837	142	10,100
46	Chiron Corporation	USA	1,723	431	5,400
47	Watson Pharmaceuticals	USA	1,641	134	3,851
48	H. Lundbeck	Denmark	1,625	296	5,155
49	Sumitomo Chemical Co.	Japan	1,622	239	20,195
50	Tanabe Seiyaku Co.	Japan	1,509	264	4,517

Source: Top 50 pharmaceutical companies, MedAdNews, September 2005