HARVARD | BUSINESS | SCHOOL

The Organizational and Geographic Drivers of Absorptive Capacity: An Empirical Analysis of Pharmaceutical R&D Laboratories

Francesca Lazzeri
Gary P. Pisano

Working Paper

14-098

April 10, 2014

The Organizational and Geographic Drivers of Absorptive Capacity: An Empirical Analysis of Pharmaceutical R&D Laboratories

Francesca Lazzeri[*]
Gary P. Pisano[†]
April 2014

Corresponding Author:
Gary Pisano
Harvard Business School
Morgan Hall 417
Boston, MA 02163
gpisano@hbs.edu

[*] Post Doctoral Research Associate, Harvard Business School
[†] Harry E. Figgie, Jr. Professor of Business Administration, Harvard Business School

Acknowledgements

We would like to thank James Zeitler, Emily Codega, and Patrick Clapp of Harvard Business School's Baker Research Services for their assistance in collecting and analyzing the data used in this study. We are grateful to William Simpson, Andrea Piccaluga, Alberto Di Minin, Giovanni Dosi for helpful input, and to participants in the Harvard Business School Technology and Operations Management seminar for comments. We gratefully acknowledge the financial support of the Harvard Business School. Any remaining errors are our own.

Abstract

Scholars and practitioners alike now recognize that a firm's capacity to assimilate and use know-how from external sources—what Cohen and Levinthal (1990) called "absorptive capacity"—plays a central role in innovation performance. In recent years, a common strategy pursued by companies to increase their absorptive capacity has been to locate new R&D facilities in close geographic proximity to technology "hotspots" like Cambridge, Massachusetts or the San Francisco Bay Area. Such a strategy is predicated on the assumption that geographic proximity facilitates absorption. Unfortunately, more than two decades after the publication of Cohen and Levinthal's landmark piece on absorptive capacity, precious little is known about how different organizational strategies and managerial practices—including location choices—actually impact a firm's ability to exploit external sources of know-how. A key barrier to empirical progress on this front has been a lack of direct measures of absorption. In this paper, we develop a novel measure of absorptive capacity that attempts to directly track the influence of external sources of know-how on the internal R&D activities on individual laboratories. We then use this measure to examine laboratory level differences in absorptive capacity and the degree to which a lab's geographic proximity to a given knowledge base influences its absorptive capacity. To identify patterns of absorption, we exploit a quasi-natural experiment that has occurred in the pharmaceutical industry over the past two decades. Since 1989, a number of major pharmaceutical companies (Merck, Novartis, Pfizer, etc.) have chosen to locate new laboratories in one or more major life science hotspots (Massachusetts, the San Francisco Bay Area, and San Diego County). Because these are *de novo* green-field labs, we have an unusual opportunity to study how the capabilities of the lab evolved over time, and whether those capabilities were influenced by the technological activities of the surrounding local scientific and technological ecosystems. Our sample includes 39R&D laboratories (at varying degrees of distance from three major life sciences hotspots—Massachusetts, San Diego County, and the San Francisco Bay Area). Our findings indicate that geographic proximity is a significant predictor of how much know-how a lab absorbs from a given hotspot. The importance of geographic proximity is also shown to be increasing over time. However, our results also show significant residual variance at both the individual laboratory and company levels, suggesting an important role of managerial practices and policies in driving absorption. The latter finding was consistent with our field interviews of R&D executives from laboratories involved in our study. The study provides further evidence of the geographically bounded nature of knowledge.

1. Introduction

The importance of external networks of knowledge for innovation is now well recognized by both scholars and practitioners (e.g., Mowery 1988; Cohen and Levinthal 1990; Galambos and Sewell 1995; Powell 1998; Owen-Smith and Powell 2004; Arora et al. 2001). Firms' ability to exploit such external source of know-how, which is what Cohen and Levinthal (1990) called "absorptive capacity"—is believed to be an important determinant of overall innovation performance. An increasingly common strategy to enhance absorptive capacity is to locate R&D laboratories in close geographical proximity to dense knowledge networks. In a variety of knowledge intensive industries such as software, electronics, and life sciences, established competitors are opening new R&D labs in such technology "hotspots" as the San Francisco Bay Area and Cambridge, Massachusetts. Unfortunately, more than two decades after the publication of Cohen and Levinthal's landmark piece on absorptive capacity, precious little is known about how different organizational strategies and managerial practices—including location choices—actually impact a firm's ability to exploit external sources of know-how.

A major barrier to progress has been a lack of direct measures of absorption. To date, most empirical studies of absorptive capacity have relied on broad proxies, such as the number of patents, academic publications, investments in R&D employees, R&D intensity (Ahuja and Katila, 2001; Belderbos et al., 2004; Cockburn and Henderson, 1998; Liu and White, 1997; Meeus et al., 2001; Mowery et al., 1996; Oltra and Flor, 2003; Stock et al., 2001; Tsai, 2001; Van Den Bosch et al., 1999; Veugelers, 1997). Further complicating matters is the fact that knowledge flows are a two-way street; any given organization is typically both a "spiller" and an "absorber" of know-how. As a result, it is difficult to ascertain whether a given technological competence of the firm was generated internally (and spilled out to others) or was instigated by absorbing externally generated know-how.

In this paper, we develop a novel measure of absorptive capacity that attempts to directly track the influence of external sources of know-how on the internal R&D activities of individual laboratories. We then use this measure to examine laboratory level differences in absorptive capacity and the degree to which a lab's geography proximity to a given knowledge base influences its absorptive capacity. To identify patterns of absorption, we exploit a quasi-natural experiment that has occurred in the pharmaceutical industry over the past two decades. Since 1989, a number of major pharmaceutical companies (Merck, Novartis, Pfizer, etc.) have chosen to locate new laboratories in one or more major life science hotspots (Massachusetts, the San Francisco Bay Area, and San Diego Country). Because these are *de novo* labs, we have an unusual opportunity to study how the capabilities of the labs evolved over time, and whether those capabilities were influenced by the technological activities of the surrounding local scientific and technological ecosystems. Because a number of laboratories still operate in legacy locations (with varying distances from the

hotspots), we have ample variance in our sample to examine the relationship between geographic proximity and absorption.

Our concept of absorption follows the general idea that absorption means the flow of some matter (knowledge in this case) across some membrane (a firm boundary in this case). Absorption can thus be measured by comparing the state of an external environment with the state of an internal environment. We follow this general approach by first constructing a detailed characterization of the know-how available in the external environment—what we call the region's "technological fingerprint"—using detailed patent data. We performed this analysis for three well-established biotechnology hotspots— Massachusetts, San Francisco Bay Area, and San Diego County. We use a similar method to characterize the distribution of technical competences inside 39 pharmaceutical R&D laboratories (some of which are located inside hotspots and some of which are not). By comparing the fingerprints of each lab with the fingerprints of different regions, we believe we have directly measured the extent to which any given lab is absorbing know-how from a given local external environment. The analysis covers the period 1980-2012.

Our analysis sheds light on three sets of questions. First, how significant are laboratory and firm level differences in absorption? As far as we know, ours is the first study to measure variance in absorptive capacity directly. Second, how is absorptive capacity influenced by geographical proximity to a hotspot? Finally, what other managerial practices may influence absorptive capacity? Our results also show that the benefits of proximity to absorption have been increasing over time. Such a finding completely contradicts the popularly accepted view that scientific and technological knowledge networks are global, and that in a flat world distance is irrelevant. Our work further confirms prior research highlighting the geographically bounded nature of knowledge networks (Baptista and Swann, 1998; Baptista, 2000; Krugman, 1991; Maskell, 2001; Pinch et al. 2003; Porter, 1990). Our results also show that while proximity helps (as expected), there is significant residual variation at both the laboratory and firm levels. This finding, along with our field interviews with R&D managers, suggests that absorptive capacity is rooted in firm-specific managerial and organizational practices.

2. Theoretical Background

An apparent paradox in the location of knowledge-based activity is that despite the intangible nature of new ideas and their potential to diffuse widely, companies within industries often cluster geographically. Previous research indicates that the benefits of spatial proximity for technological innovation do not spring from random spillovers between unconnected parties, but through specific mechanisms that facilitate knowledge exchange within a region: social and professional contacts, informal communication, and local labor market turnover (Almeida and Kogut 1999, Owen-Smith and Powell 2004; Saxenien 1994; Breschi and Lissoni 2001). Proximity allows for face-to-face interaction

and transfer of tacit knowledge. Audretsch and Stephan (1996) show that localization between biotechnology firms and outside scientists is strongest for relationships that entail direct knowledge transfer.

Of course, the observation that firms within the same industry tend to cluster geographically is not new. Almost a century ago, Alfred Marshall (1920) observed such clustering in Britain--what he called industrial districts. Marshall hypothesized that firms from the same industry cluster geographically to exploit common labor pools, common infrastructure and suppliers, and spillovers of knowledge. Relatively more recent work on industrial districts (e.g. Best, 1990; Piore and Sabel, 1984; Porter, 1990; Saxenian, 1994; Feser 1998) has emphasized the importance of spatial proximity and interfirm networks.

Technology spillovers occur when a firm receives economic benefit from another firm's R&D activity without incurring the same cost. Jaffe (1986, 1989) is one of the first researchers to quantify the extent of spillovers within geographies. Using a modified knowledge production function (Griliches, 1979) with a spatial component that measures the importance of geographic proximity for university and industry research, Jaffe found evidence of geographically mediated spillovers from university research, especially in drugs, chemicals, and electronics. Jaffe's results suggest that when companies locate their R&D centers in a technological cluster, they may benefit from the knowledge of other firms, as well as unintentionally facilitating the learning and transfer process of some of their own corporate knowledge to other companies in the cluster.

Spillovers, however, can only occur if firms have the capacity to absorb know-how from external sources. In the economics literature on spillovers, it is assumed that all firms benefit equally from knowledge spillover flows. However, in the management literature, Cohen and Levinthal's theory of absorptive capacity predicts that the capacity to absorb knowledge from the external environment may differ among firms (Cohen & Levinthal 1989, 1990, 1994). Moreover, Cohen and Levinthal's works suggest that the ability of companies to absorb external know-how could be rooted in such factors as the firm's knowledge stock.

In their analytical model, Cohen and Levinthal (1990) use absorptive capacity as a variable to explain the effect of appropriability conditions and technological opportunity on R&D intensity. In this sense, they use absorptive capacity as a conceptual tool to determine the incentives for R&D investment, but do not establish a means of measuring absorptive capacity directly. Cohen and Levinthal (1990) use their model to conclude that an increase in external information (technological opportunity) would automatically lead to an increase in the incentives to build absorptive capacity. In many ways, firms that choose to locate R&D laboratories in a region of high technological opportunity (a hotspot) are behaving in a manner highly consistent with the predictions of the Cohen and Levinthal model.

Since Cohen and Levinthal's seminal work, many empirical and theoretical studies have explored the concept of absorptive capacity from the perspective of different analytical units and modeling strategies. Lane et al. (2006) observe that most researchers typically measure absorptive capacity with simple R&D proxies, ignoring the variety of its dimensions and their implications for different organizational outcomes. The use of these proxies may have contributed to conflicting and ambiguous findings about the nature and contributions of absorptive capacity. For example, R&D spending is not the only source of absorptive capacity since employee skills, organizational memory, prior organizational experiments and experiences and even (as we examine later) location may contribute significantly to a firm's overall absorptive capacity.

Several authors have attempted to refine absorptive capacity definitions (Dyer and Singh, 1998; Lane and Lubatkin, 1998; Van den Bosch et al., 1999; Zahra and George, 2002; Lane et al., 2006), by considering the multiple dimensions of absorptions (knowledge acquisition, assimilation, transformation, and exploitation). Unfortunately, because different authors have tended to focus on different dimension of absorption, comparing findings across studies is difficult. For example, Ahuja and Katila (2001) focus on the "acquisition" dimension of absorptive capacity, described as the generation of insights from various sources, and use the number of patents issued by companies to study technological acquisition and firm performance; other authors (Meeus et al., 2001) use R&D-intensity as a proxy to explore the "assimilation" dimension of absorptive capacity, that they explain as a dissemination of knowledge within the organization; other authors focus on the "transformation" (combination of existing knowledge and newly generated knowledge) and "exploitation" (use of transformed knowledge for product development and for the benefit of the overall organization) dimensions, using the number of employees with Ph.D's advanced degrees as proxies (e.g. Veugelers, 1997; Muscio, 2007).

Despite these efforts, we still lack a common means to measure directly the flow of knowledge from the external environment into the firm's internal environment. Without such a measure, we cannot characterize the variance across organizations in absorptive capacity and cannot even begin to probe the organizational and other factors that may drive that variance. The present study tries to solve the absorptive capacity measurement constraint by developing a direct measure of absorption. While we have used the pharmaceutical industry as an empirical context, we believe the method is general enough to apply to other technological and industrial contexts.

3. Overview of data and methodology

Pharmaceutical industry is an ideal context for our empirical analysis for a number of reasons. First, over the past three decades, sweeping changes in the scientific underpinnings of pharmaceutical R&D have created the kind of fertile external "technological opportunity" identified by Cohen and Levinthal (1990) as an incentive for incumbent firms to deepen their absorptive capacity. Because these changes—generally grouped under the broad heading of

biotechnology—originated largely in universities and entrepreneurial companies (and outside the boundaries of established pharmaceutical companies), they created a significant external source of relevant know-how (Pisano 1990, 1996). Secondly, the geographic sources of biotechnology have been relatively concentrated in a number of distinct regional hotspots like Massachusetts, the San Francisco Bay Area, and San Diego County (Zucker et al., 1994). Finally, because these hotspots emerged in places not historically associated with pharmaceutical R&D, decisions by some incumbent firms to locate laboratories their provides a quasi-natural experiment to examine the impact of distance on absorption.

To gain a better understanding of the factors influencing laboratory location and the processes by which firms seek to absorb know-how from external sources, we conducted interviews with R&D executives at eight major pharmaceutical companies. All of these companies had multiple laboratories. Five had at least one laboratory in one of our three designated hotspots. Three others did not yet have a laboratory in a hotspot. We interviewed a total of 16 R&D executives across these companies. Interviews were semi-structured and lasted an average of 45 minutes. Information from these interviews helped us design our statistical model, and as we explain later, helped us interpret our findings.

One critical piece of information we gleaned from these interviews concerned how firms decided to put labs in specific locations. All of the firms that had built new labs in hotspots reported (consistently) that access to talent was a primary motivator. They also believed that being geographically close to important academic institutions and entrepreneurial firms was important to gain access to leading-edge science. We asked in every interview whether the decision to establish a lab in a specific hotspot (say, Massachusetts) was motivated by a prior interest in a *particular technology* that they believed was uniquely available in that hotspot. For instance, was a firm with a particular interest in kinase inhibitors locating their new R&D laboratories in Massachusetts because they believed Massachusetts was the best place to access kinase inhibitor knowledge? Our interviewees were universal in reporting that consideration of specific technology fields was not a driver of their decision. Because of the rapid pace of technological evolution in life sciences (hot fields change relatively quickly), they reported it would be very difficult to identify in advance which specific streams of technology might be most attractive in any given place in the future. Thus, we feel confident that while firms were locating labs in hotspots to access general life sciences and biomedical know-how, the choices of lab location was not endogenously determined by specific technological interests.

Our quantitative method involved the following steps. First, we created detailed characterizations of the knowledge environment in each of the three major biotechnology ecosystems in our study (Massachusetts, the San Francisco Bay Area, and San Diego County). We refer to these profiles as technological fingerprints. We can think of the technological fingerprint of each hotspot as representing the pool of knowledge potentially available in that location.

Second, for each of the 39 laboratories in our study, we created a similar profile or fingerprint. A lab's fingerprint can be viewed as representing the know-how generated by the firm. The fingerprints of both the hotspots and each laboratory are calculated over the period 1980-2012. Third, we calculate the absorptive capacity of individual laboratories by comparing how closely each lab's internal R&D activities track the evolution of technology in the external environment. In our method, absorptive capacity is a relative construct in that it can only be measured relative to some target knowledge base. Thus, it makes no sense to talk about, say, the absorptive capacity of Novartis' research laboratory in Cambridge Massachusetts in the absence of a reference point. We can only define that lab's absorptive capacity relative to a specific body of knowledge from a specific region, such as Massachusetts or California. Fourth, using our measure of absorptive capacity, we use regression analysis to examine the impact of distance and other factors on absorption.

3.1 Mapping Regional Technological Fingerprints

The starting point in our analysis is to characterize the distribution of knowledge in each of our three targeted hotspots: Massachusetts, San Francisco Bay Area, and San Diego County. As noted earlier, we chose these three hotspots because they have each developed fertile biotechnology ecosystems. In addition, each has become the home of at least one new R&D laboratory of an incumbent pharmaceutical company. To create a technological fingerprint for each region we used detailed data on patent granted between 1980 and 2012. We chose patents as our raw data for two reasons. First, patents are a critical resource in the pharmaceutical and biotechnology industry. Academic institutions, young firms, and established firms all seek to protect their intellectual capital vigorously by patenting. Thus, patents provide a reasonable picture of the R&D activities of firms and other members of the ecosystem. Second, the patent classification system, while certainly not perfect, provides a standardized way for us to characterize the scientific and technological content of the know-how created within regions and by labs. We utilized patents granted (as opposed to patents applied for) because the patent examination process provides a standardized degree of quality control in terms of novelty and classification.

We obtained patent data from the USPTO via the Thomson Reuters Westlaw database. We take advantage of the US Patent Classification system to identify the technological and scientific content of each patent in our dataset. The U.S. Patent Classification System is a system for organizing all U.S. patent documents and many other technical documents into relatively small collections based on common subject matter. Each subject matter division in the USPCS includes a major component called a class and a minor component called a subclass. A class generally delineates one technology from another. Subclasses delineate processes, structural features, and functional features of the subject matter encompassed within the scope of a class. A class/subclass pair of identifiers uniquely identifies a subclass within a class: for example, in our case the technological sub-class 424/9.81 refers to any antigenic substance

or allergen, that "is applied to or injected into a subject in order to determine whether the subject is allergic or hypersensitive to the agent as indicated by a visible change on the skin (i.e., redness, swelling, etc.)" (http://www.uspto.gov/web/patents/classification).

The nine specific USPTO classes included in our data sets were 424, 435, 514, 530, 536, 800, 930, 935, and 436, that have been defined as biotechnology classes in the literature (Adelman and De Angelis, 2007; Johnson, 2009; USPTO Technology Assessment and Forecast Program, 1998). Specifically, the definition includes U.S. patent sub-classes 47/1.1-47/1.4, 47/57.6-47758, 424/9.1-424/9.2, 424/9.34-424/9.81, 424/85.1-424/94.67, 424/130.1-424/283.1, 424/520-424/583, 424/800-424/832, 435/1.1-435/7.95, 435/40.5-435/261, 435/317.1-435/975, 436/500-436/829, 514/2-514/22, 514/44, 514/783, 530/300-530/427, 530/800-530/868, 536/1.11-536/23.74, 536/25.1-536/25.2, 800, 930, 935.

The patents in our sample were granted between 01/01/1980 and 12/31/2012. Our patent database consists of 78,539 patents (48,039 patents originating from the San Francisco Bay Area, 18,020 from Massachusetts, and 11,480 from San Diego County). Using inventor address data contained on each patent, we identified all patents originating from one of our three target hotspots (we classified a patent as belonging to a specific region if at least an inventor's address was from that region). Specifically we looked at the inventors' zip codes. We defined the San Francisco Bay Area as including the following nine counties: Alameda, Contra Costa, Marin, Napa, San Francisco, San Mateo, Santa Clara, Solano, and Sonoma. We defined Massachusetts as including the following fourteen counties: Barnstable, Berkshire, Bristol, Dukes, Essex, Franklin, Hampden, Hampshire, Middlesex, Nantucket, Norfolk, Plymouth, Suffolk, and Worcester. San Diego County is a single county. We selected these specific geographic boundaries based on our knowledge of the geographic concentration of life sciences research in each of these three regions.

In order to analyze the technological fingerprints of each hotspot, we looked at the distribution patents over all above identified patent sub-classes. Each hotspot's technological fingerprint can be characterized by a vector:

$$F_{it} = (f_{i1}, f_{i2}, \ldots, f_{iK})$$

where *Fit* refers to the number of patents in sub-class i for each year.

3.2 Mapping Technological Fingerprints of R&D Laboratories

Our sample of R&D laboratories was constructed as follows. We first identified all of the laboratories that incumbent pharmaceutical companies had established in any one of our three target hotspots. We included only those laboratories that were *de novo* (green-field) operations, and excluded any labs acquired through acquisition. We chose to focus only on *de novo* laboratories because they enable us to track the evolution of a lab's know-how from a clearly defined initial starting point (acquired labs come with prior absorption history). This process yielded 10 laboratories in total (6 in Massachusetts, 3 in San Diego County, and 1 in San Francisco Bay Area). Eight incumbent pharmaceutical companies were represented in this sub-sample (Abbott, Amgen, Astra Zeneca, Johnson & Johnson, Merck, Novartis

and Pfizer). To create variance along our primary independent variable of interest, we then identified the other major legacy R&D laboratories of these companies located inside the US. This yielded an additional 18 laboratories for our sample. To create additional variance in the sample of companies, we also included the US laboratories of several companies in the industry that had no *de novo* hotspot laboratories as of 2012 (Bristol Myers Squib, GlaxoSmithKline, Eli Lilly, Sanofi, and Boehringer Ingleheim). This yielded an additional 11 laboratories for our sample. Our total sample thus consisted of 39 laboratories (10 of which were located inside one of our designated hotspots, 28 of which were located elsewhere). Table 1 provides a list of laboratories (by location and corporate affiliation) in our study.

[Table 1 Here]

We then created a technological fingerprint for each of the 39 labs in the study using the same method as used for mapping the technological fingerprints of the three hotspots. Using a combination of the company assignee and the inventors' addresses (listed on the patent), we identified the patents most likely to have originated from each of our sample laboratories. Patents with multiple inventors whose addresses spanned multiple potential laboratories were assigned to multiple laboratories (e.g. a Merck patent with an inventor in Boston and an inventor in New Jersey would be attributed to both Merck's New Jersey laboratory and to its Massachusetts laboratory). For each laboratory, we then constructed a vector for each year composed of the distribution of patents across all subclasses.

3.3 Measuring the technological distance

In this study the basic data characterizing the R&D labs and regional hotspot' technological fingerprints were patent subclasses-based distribution vectors. Using these vectors, we calculated the technological distance between each of our 39 sample laboratories and each of our three hotspots. A large technological distance indicates a low concordance between the patent portfolio of the lab and the patent portfolio of a given region, and thus a low level of absorption. Conversely, short technological distances are indicative of high absorption rates. Technological distance is our measure of absorption and will be the dependent variable in our regression analyses below.

There are various measures to express (dis)similarity between pairs of objects, such as the Angular, Canberra or Mahalanobis distance (Kaufman and Rousseeuw, 2005). For our purpose, we have decided to use the Euclidean distance, defined as the square root of the sum of the squared differences in the variables' values and it corresponds to the length of the line that connects two points. The Euclidean distance is the most commonly used type of distance when it comes to analyzing ratio or interval-scaled data and may be seen as a special case of the Mahalanobis distance with equal variances of the variables. One important feature of the Euclidean distance is that it is invariant over the choice of origin and orientation of coordinate axes. The choice of using Euclidean distance in hierarchical clustering

procedure is really a choice between using the full-covariance of clusters or ignoring them: when we decide to use Euclidean distance, we assume that the clusters have identity co-variances.

Since we are assuming that the scale steps between our ordinal data are equidistant and we are in a non-spatial space, but in a technological one, we considered the Euclidean distance the most accurate measure to identify the technological distance between the labs and the three hotspots. Other studies have used the Euclidean distance to measure the distance between data points (Andrews and Currim, 2003; Arabie and Hubert, 1994; Bishop, 2006; Jaskowiak et al., 2013). Those studies usually used the Euclidean measure to compute distances from row data and not from standardized data: indeed, the Euclidean measure has the advantage that the distance between any two objects is not affected by the addition of new objects to the analysis, which may be outliers.

Note, in measuring the distance between any given laboratory and a given hotspot, we excluded the patents of that lab from the vector of the hotspot. For instance, in calculating the distance between the patent vector of Novartis-Cambridge and Massachusetts hotspot, we excluded the Novartis patents from the vector characterizing the Massachusetts. While each lab (even the largest) typically only represents a small fraction of the total population of patents in any given region, our process ensures that we have avoided the obvious bias that would be created by counted by measuring the same patents in both the lab and the external environment.

4. Characterizing the Variance in Absorption at the Laboratory Level

Using the above method, we can develop a picture of how well each individual laboratory in our sample absorbs know-how from each of the three geographic hotspots in our study. One of the challenges when using any new metric is calibration and validation. Does our measure of technological distance between the labs patent profile and that of a given hotspot really depict absorption? One way to validate our measure is to consider labs that a priori we would expect to have very high rates of absorption of know-how from each geographic hotspot. For each hotspot, we chose one particular lab (outside of our sample) that due to its long history in the region and deep local routes should have very high rates of absorption. We also sought biotechnology firms that have remained by and large independent for the major chunk of our sample period (that is, we did not want firms that had operated as subsidiaries of major pharmaceutical companies). Using these criteria, we identified a validation laboratory for each of our three hotspots. For San Francisco, we picked biotechnology pioneer Genentech as our validation laboratory. Founded in 1976, Genentech has been a major R&D presence in biotechnology and is by far the largest biotechnology firm in the San Francisco Bay Area. It was an independent firm until 2010 when the Swiss pharmaceutical company Roche acquired it. For Massachusetts, we picked Biogen (later renamed Biogen Idec after a merger). Biogen was founded in 1978 by Harvard molecular biologist Walter Gilbert. It was the first biotechnology firm in Massachusetts, and is the largest independent biotechnology firm in the region. For San Diego County, we chose Amylin, founded in 1987 and

independent until 2012 when Bristol-Myers Squibb acquired it. Our expectation is that if our measure is reasonable, all three of these firms' patent portfolios should demonstrate short technological distance (i.e. high absorption) relative to their home hotspots. There is no perfect way to validate a measure, and we consider our approach here as a simple sanity check.

We calculate our distance measures for every single year a laboratory is in our sample. We are initially interested in understanding the degree of variance in absorption across laboratories. As a means simply to illustrate the degree of variance (a more rigorous analysis is conducted later), we provide the values for technological distance between each laboratory and each hotspot for the final year in our sample (2012). Note, we calculated the technological distance of each R&D lab from the three hotspots for each year, but for purposes of graphical illustrative are only showing the final year. We also include in our graphs the "validation" laboratory for each hotspot. These are shown in Figures 1, 2, and 3.

[Figures 1, 2, and 3 Here]

The three figures illustrate that for each of our hotspots the range of variance in absorption across labs is quite wide. As hoped for, our validation labs (Genentech for San Francisco Bay Area, Amylin for San Diego County, and Biogen for Massachusetts) demonstrate very short technological distances from each of their respective hotspots. A cursory examination of the variance by laboratory suggests that geographically proximate laboratories have higher absorption rates (shorter technological distances) from their local hotspots than more distant laboratories. A more systematic analysis of the impact of geographic location and other factors on absorption is performed in our regression analyses below.

5. Model Specification

Our primary interest lies in understanding the extent to which a lab's geographic distance from a given hotspot influences that lab's ability to absorb know-how from that hotspot. Our primary independent variable of interest is the geographic distance between the lab and the hotspot. Using route planning software program (http://www.distancefromto.net) we computed the geographical distance in miles from each lab and the respective hotspot. For this type of analysis, a reference point (ground zero) needs to be chosen. Our choice of ground zero reference point was guided by information on the general accepted epicenter of scientific and entrepreneurial activity in the life sciences for each hotspot. For the Massachusetts hotspot, we used Kendal Square (home to MIT, the Whitehead Institute, the Broad, and a large number of start-up companies) as ground zero. For the San Diego County hotspot, we used the Scripps Research Institute in La Jolla, CA as ground zero. Scripps is one of the premier research institutes in life sciences and La Jolla is home to a significant fraction of biotechnology firms. For the San Francisco Bay Area, our

ground-zero reference point was Genentech in South San Francisco. This reference point was chosen for several reasons. First, Genentech, being one of the pioneers in the biotechnology field, is a considerable source of scientific and technological know-how. Second, South San Francisco, and more specifically the industrial park off Highway 101 (Grand Avenue Exit) that is home to Genentech is also home to many young biotechnology firms. And finally South San Francisco is approximately midway between two major academic research centers—University of California, San Francisco and Stanford University-- that have played a prominent role in life sciences research.

We looked at the impact of location on absorption in three different ways. We first used a simple log transformation of geographical distance values: the log transformation can be used to make highly skewed distributions less skewed. This can be valuable both for making patterns in the data more interpretable and for helping to meet the assumptions of inferential statistics. After these transformations, the variables satisfy all basic assumptions of the regression models presented hereafter.

To explore whether the effect of distance on absorption was highly non-linear, we also specified a version of the model with dummy variables indicating different ranges of distance between each lab and each hotspot. We created 5 different distance dummy variables named respectively: Distance 1 (dummy variable equals to 1 if the distance between the lab and the hotspot is < or = to 5 miles; 0 if > than 5 miles); Distance 2 (dummy variable equals to 1 if a lab is located between 6 and 20 miles from the hotspot; 0 if > than 20 miles); Distance 3 (dummy variable equals to 1 if a lab is located between 20 and 50 miles from the hotspot; 0 if > than 50 miles); Distance 4 (dummy variable equals to 1 if a lab is located between 50 and 150 miles from the hotspot; 0 if > than 150 miles). Distance 5 (dummy variable equals to 1 if the distance between the lab and the hotspot is > than 150 miles; 0 if < than 150 miles).

In a third specification, we introduced two location dummy variables indicating whether the laboratory was located in Massachusetts or San Diego County (because there was only 1 *de novo* hotspot laboratory in San Francisco Bay Area, we excluded that category, and it is included with the residual "all others" category). This locator dummy can tell us two things. First, it provides another discontinuous means of assessing the impact of distance of absorption. For the Massachusetts model, the dummy variable "Massachusetts" is essentially an indicator of a lab being local (without differentiating distance). A similar logic applies for the San Diego model. Second, the dummy indicator also tells us if there is something about a firm's absolute location (not relative to a hotspot, but simply its location) that might impact its absorptive capacity. For instance, using this dummy variable, we can understand whether labs in, say, Massachusetts have better absorptive capacity than labs in other locations (irrespective of whether they are absorbing know-how from San Diego, San Francisco Bay Area, or Massachusetts). Is there something about "being in a hotspot" that makes a lab better able to absorb know-how from any hotspot?

5.1 Additional Variables

Our data set includes 32 years (from 1980 to 2012) of observations. Our time series analysis accounts for the fact that data points taken over time may have an internal structure (such as autocorrelation, trend or variation) that may affect the technological distance dependent variable. However, our panel data is unbalanced, since the number of observations per time period varies: the number of labs per year is not always the same because some labs in our sample have been opened in different years. To control for the influence of time, we included a continuous variable indicating the year of the observation.

Our field interviews suggested other factors could influence absorption at both the laboratory and company level. Firms in our sample had varying policies and norms with respect to publishing, collaborating with academic investigators, and attending academic seminars. Some laboratory management explicitly sought to foster ties to the external scientific community by establishing their own seminar series and scientific symposia (open to scientific community), creating joint post-doctoral programs with local universities, and hiring scientists who were well connected to the local scientific community). At this stage in our research, we did not have access to enough firms to systematically examine how individual policies might impact absorption (this is the subject of future research). However, together, the information obtained from these interviews strongly suggests that we include both firm and laboratory level fixed effects to capture the potential impact of these policies on absorption.

We also learned that firms have different approaches to sharing knowledge between internal laboratories within their own network. Some firms, for instance, explicitly seek to disseminate knowledge obtained from any of their labs as broadly as possible throughout their laboratory networks. Others chose a more "focused factory" type approach in which labs in different locations essentially specialized in different fields of research (usually by therapeutic or disease field). The impact of these differences should also be captured with firm fixed effects.

As an alternative to firm fixed effects, we included a variable that captured the measured absorptive capacity of other labs in each firm's corporate network (sister laboratories). A sister effect would be suggested if the absorptive capacity of a given lab can be predicted by the absorptive capacity of other labs in the company's network. The impact of different network strategies on the absorption of any given laboratory should be captured by this variable.

To control for possible scale effects at the laboratory level, we included a variable measuring the total number of patents issued by each lab in a certain year ("Patents"). Ideally, scale effects would be captured by total spending or headcount at each laboratory, but such highly proprietary data were not available to us. Given prior research that shows a link between total R&D spending and total patent output, we believe our patent variable is a suitable proxy for laboratory scale. Finally, we created an interactive variable - Time*Log_Distance-Miles – in order to test if the impact of geographical distance on technological distance is increasing or decreasing over time.

We estimated the following model:

$$\text{TechDist}_{ijt} = \beta_{0j} + \beta_{1j}(\text{Log\_Distance-Miles})_{ijt} + \beta_{2j}(\text{DistanceDummies})_{ijt} + \beta_{3j}(\text{LocationDummies})_{ijt} + \beta_{4j}(\text{Time})_{ijt} + \\ + \beta_{5j}(\text{FirmFixedEffects})_{ijt} + \beta_{6j}(\text{LabFixedEffects})_{ijt} + \beta_{7j}(\text{Patents})_{ijt} + \beta_{8j}(\text{Time} * \text{Log\_Distance-Miles})_{ijt} + \varepsilon_{ijt}$$

where i indexes for the lab, j indexes for the firm, and t indexes for time. Variable descriptions are provided in Table 2.

Our analysis included three regions (San Diego County, Massachusetts, and San Francisco Bay Area). Rather than interacting location dummies against all the variables in the equation, we ran a separate analysis for each region. While the sample of 39 labs stays the same, values of variables like distance change depending on the target region of the model. For instance, in the Massachusetts model, the Novartis lab in Cambridge, MA would be considered geographically quite proximate, whereas that same lab in the San Diego model is now relatively distant geographically. Since our data have a nested structure (we have three levels of observation in our model: the company level, the lab level, and the company's network level), we engaged in a multilevel regression. With this type of data, classic methods, such as OLS regression, would not produce correct standard errors (Bryk and Raudenbush, 1992; Hox, 2002). Therefore, multilevel models need to be used as they take correlated errors into consideration. Multilevel regression models provide estimates of higher (firm) level variables on lower (lab) level outcomes, while accounting for the non-independence of observations within labs. A simple variant is the random intercept model. Such models treat differences between labs as a source of variance in the intercept of the regression equation (Snijders and Bosker, 1999).

6. Empirical results

The results of our analyses are shown in Tables 3, 4 and 5. To examine if our data call for multilevel modeling, we conducted various a priori tests as recommended by Snijders and Bosker (1999). The interclass correlation coefficient was 0.17 and positive, while one-way analysis of variance revealed significant differences in the technological distances of labs to the hotspots ($F = 1.84$, $p < 0.001$). This implies that ordinary least squares estimates would provide inaccurate standard errors and false tests of significance (Snijders and Bosker, 1999). As multilevel regression uses maximum-likelihood estimators, model fit is assessed by comparing deviance measures of subsequent models: a decrease of the deviance measure (Δdev) is related to Δdf (degrees of freedom) and tested against a $\chi 2$-distribution. Model 0 gives an initial deviance value of 3309.61. In order to be able to compare the results between each hotspot, we decided to run the regression analysis for each region separately.

6.1 The Massachusetts regression results

The table below (Table 3) presents the results from the multilevel regression models of technological distance in the Massachusetts hotspot. Pseudo-R2 is computed according to the guidelines of Snijders and Bosker (1999).

[Table 3 Here]

Model I examines the relationship between technological distance and geographical distance. The idea that geographic proximity facilitates absorption is supported by the positive and statistically significant coefficient on the Distance variable ($b = 0.14$, $p < 0.01$); technological distance increases with geographical distance, meaning that laboratories geographically close to the Massachusetts hotspot have shorter technological distances (higher absorption).

Our dummy indicators of distance provide some insight about the thresholds at which distance may become important. Here, our dummy variable indicating a laboratory location within 5 miles of Kendall Square (Distance 1) is negative and significant ($b = -0.46$*, $p < 0.01$), while the other distance threshold indicators (Distance 2, 3, 4) are positive and significant. These results indicate that labs within a 5-mile radius of Kendall Square have higher rates of absorption than labs outside that radius. This is quite a tight geographic window, suggesting that the benefits of distance for absorption dissipate rapidly over distance.

The Massachusetts Location dummy coefficient is significant and negative ($b = -0.45$, $p < 0.001$), providing further confirmation of the benefits of distance on absorption. In later models exploring the other regions, we can compare how this dummy changes to see whether there is an'intrinsic' advantage of being a Massachusetts laboratory vis-à-vis any hotspot location.

The variable Time is significant and negative ($b = -0.28$, $p < 0.001$), indicating that over time all labs are getting closer to the know-how generated by hotspots (regardless of their location). While prior studies have demonstrated that knowledge can be transferred more easily within a firm than across firms (Darr et al., 1995), our results suggest that over time capability transfers and knowledge diffusion effects among networks within the same industry are particularly likely to increase. Firm and lab fixed effects were significant ($p < 0.01$) indicating the potential importance of both corporate level and firm level management practices and policies in shaping absorptive capacity. In model III we added the Patent variable and the Time*Log_Distance-Miles variable. Scale effects were not significant. The coefficient on the Time*Log_Distance-Miles variable was both positive and significant ($b= 0.14$, $p < 0.01$), indicating that over time, the impact of distance on absorption is getting greater over time.

Finally, Model IV examines an alternative specification for the firm effects, using information from the absorption levels of other sister labs in the company networks. We call this phenomenon the sister effect. In this model we did not include the Firm Fixed Effects. The sister effect will tell us something about the diffusion of knowledge

across laboratories in the same company. If know-how diffuses rapidly from hotspot labs to other labs in the companies' network (and vice versa), the coefficient on the sister effect should be positive. A negative coefficient would suggest that firms are specializing labs by technological field, or following a geographic division of labor strategy for their labs (e.g. Massachusetts lab focus on Massachusetts know-how, San Diego labs focus on know-how from that region, etc.). Our field interviews suggested that most firms were at least trying to follow the former strategy of diffusion. Our statistical results indicate otherwise. The coefficient of the Sister Effect variable is significant but negative ($b = -1.27$, $p < 0.01$). This means that if a company has a laboratory in Massachusetts with high rates of absorption (of Massachusetts know-how), then it is more likely that its non-Massachusetts laboratories will have low rates of absorption (of Massachusetts know-how). This suggests that on average companies in our sample are following (perhaps implicitly) geographic division of labor strategies for their laboratories. It may also indicate hidden organizational barriers to the diffusion of know-how across internal corporate laboratories.

There was no significant effect of scale on absorption.

6.2 San Diego County and San Francisco Bay Area Results

We repeated the above analysis for both San Diego County and the San Francisco Bay Area using the same sample of laboratories. The results are shown in Table 6 (San Diego) and Table 7 (San Francisco Bay Area).

[Table 4 and Table 5 Here]

The results for both the San Diego County and San Francisco Bay Area hotspots are comparable to the Massachusetts results. For both of those (as well as Massachusetts), there is strong evidence that geographic proximity and absorptive capacity are correlated. Being close helps. The only notable difference between the hotspots concerns the distance threshold effects. San Diego County (like Massachusetts) had a relatively tight geographic window on geographic proximity. Labs within a 5-mile radius of Scripps had higher rates of absorption of local know-how than labs outside the 5 mile radius. However, for the San Francisco Bay Area, the geographic window was larger. The threshold for higher absorption occurs within a 50-mile radius of Genentech. One should not read too much into these differences in terms of the diffusion of know-how. They may simply reflect the different institutional topographies of each region. In both Massachusetts and San Diego, the influential research institutions are tightly clustered geographically. Within 5 miles of Kendall Square, for instance, there is MIT, the Whitehead Institute, the Broad Institute, Harvard University, Harvard Medical School, the major Harvard teaching hospitals (the Massachusetts General Hospital, the Brigham and Women's, the Beth Israel Deaconess), Boston University, and several other prominent research hospitals (Children's Hospital, the Dana Farber Cancer Institute, etc.). Within 5 miles of the Scripps Institute, one can find the University of California-San Diego, the Salk Institute, the Sanford-Burnham Medical Research Institute, and the teaching hospitals of

the University of San Diego. The institutional topography of the San Francisco Bay area is relatively more spread out. The University of California San Francisco (located in the city of San Francisco) is 37 miles from Stanford University (in Palo Alto) and approximately 17 miles from the University of California, Berkeley (Stanford and Berkeley are 39 miles apart).

All other results concerning distance, location, time, scale, and laboratory- and firm-fixed effects are the same. We again see that being in a particular hotspot helps absorption with know-how from that hotspot only. The Massachusetts hotspot labs that had an advantage in absorbing know-how from Massachusetts were at a disadvantage relative to San Diego labs in absorbing know-how from San Diego (as well as for San Francisco). Note, there is no intrinsic benefit of being located in a hotspot location in terms of absorbing know-how from other hotspots. We again see that the advantage of proximity is increasing over time for both San Diego and the San Francisco Bay Area. And, we again see that the sister effect is negative. There is, as in the case of Massachusetts, no significant impact of scale for either San Diego or San Francisco Bay Area.

8. Discussion

Our results suggest that geographic location is important factor influencing an organization's capacity to absorb know-how from external sources. The results also suggest the effects of proximity have a fairly tight threshold (it is valuable to be very close to the epicenter, but once outside that tight radius, the impact falls off considerably). Being geographically proximate to a source of know-how enhances the degree to which an organization can absorb know-how from that particular source. Labs close to Massachusetts had higher rates of absorption (shorter technological distance) of Massachusetts originating know-how than labs that were further away. The same was true for San Diego and the San Francisco Bay Area. These findings suggest that laboratory locations choices are a critical ingredient of a firm's technology strategy given the importance of external know-how to a firm's overall innovation performance. Because it is impossible for any given lab to be simultaneously close to all hotspots, location choices involve trade-offs. The lab in Kendall Square (Cambridge, Massachusetts) that has a decided advantage in absorbing know-how from the Boston area life sciences ecosystem is at a disadvantage in absorbing know-how from San Diego. The lab across the street from the Scripps Institute in La Jolla may have a distinct advantage in absorbing know-how from the San Diego scientific ecosystem, but is disadvantaged when it comes to absorbing know-how from Massachusetts. This finding may explain why several larger pharmaceutical companies (like Novartis, Merck, and Pfizer) have chosen to locate new laboratories in several hotspots. This suggests that scale at the company level, which enables a firm to afford a more geographically diverse lab network, may be an advantage in absorptive capacity and thus innovation. This is a hypothesis that should be tested in future work.

Interestingly, we found that influence of proximity increasing over time (against, consistently across all three hotspots). One possible explanation of this result is that the scientific networks inside the hotspots are becoming denser with time. As this process has occurred, and as the number of established pharmaceutical laboratories inside hotspots have increases, the liabilities of being non-local have increased. Take Massachusetts as an example. Prior to 2000, there was only 1 established pharmaceutical company laboratory in the state. By 2012, there were six, meaning that academic scientists and entrepreneurial firms had many more local choices for collaboration. This increasing density of the knowledge networks, however, is speculation, and should be subject to further research.

While location was found to be a statistically significant predictor of absorption, there were also significant laboratory and firm level fixed effects. While the literature on absorptive capacity has tended to talk about absorptive capacity as a firm level capability, our results suggest that it operates at both the firm level and organizational sub-unit (laboratory) level. In our field interviews, we certainly learned about firm level policies that might affect absorption (positively or negatively). For instance, firms in our study differed greatly in terms of intellectual property (IP) policies that might help (or hinder) outside collaboration, a key conduit for absorption of external know-how. Some firms in our study reported IP policies that heavily restricted their scientists' freedom to collaborate and publish with academic scientists. Others described policies that were more flexible with respect to sharing know-how with outsiders. These kinds of company level policies (often set by the company's legal department) would be expected to influence absorption across all the company's laboratories. Thus, our very limited sample of field interviews and our statistical results concur with the general argument of the literature that absorptive capacity is firm-specific capability. However, the significance of laboratory fixed effects also suggests that absorptive capacity varies significantly across laboratories within the same company. Thus, not all absorptive capacity "lives" at the firm level.

We found in our study that laboratory level management had significant discretion in running their laboratories, and establishing policies that might influence absorption (e.g. recruiting, prioritizing external collaboration, etc.). Our sample of within-firm laboratories was too small to make systematic comparisons of these policies (only one firm in our study had a labs in all three hotspots, e.g.). We are currently exploring the potential influence of laboratory level management policies through a separate in-depth case study.

9. Conclusion

The ability of firms to exploit know-how from external sources has long been theorized to be an important determinant of overall innovative performance (Cohen and Levinthal, 1990). Unfortunately, it has been virtually impossible to test this hypothesis, or to distill its practical implications, because of challenges measuring absorption itself. In this paper, we attempted to make progress on the task of measuring absorption, characterizing its variance, and identifying some important covariates.

Our measure of absorptive capacity has two distinguishing characteristics. First, absorption for us is a relative concept; an organization's absorptive capacity can only be measured relative to some identifiable target body of knowledge. One cannot say, for instance, that Organization A has better absorptive capacity than Organization B. It can only be said that Organization A has better absorptive capacity of body of knowledge X than Organization B. In this study, we exploited the fact that in the life sciences, we can identify distinct bodies of knowledge emanating from different geographies (or hotspots). This allowed us to examine how well a sample of organizations absorbed know-how from each of those distinct bodies of knowledge, and to systematically explore the impact of geographic distance on absorption. Second, absorption for us is mimetic. In our measure, a high level of absorption is indicated by a close matching between the distribution of patents in the laboratory's portfolio and the distribution of patents in the external environment. The more closely the labs portfolio matches the portfolio of the environment, the greater we presume absorption to be.

As with all measures, there are strengths and weaknesses of our approach. The strength of our approach is that we can clearly identify a target body of external knowledge. This approach is flexible enough to be used with non-geographic boundaries as well. For instance, if a researcher can identify ex ante the most relevant bodies of know-how in the external environment, then exactly the same fingerprint matching methodology can be utilized. It can and also should be tested outside the confines of the pharmaceutical industry to see if similar results are obtained. Clarity, however, comes at a cost. Our concept of absorption assumes that imitation is a key mechanism of the process. From our field interviews, we believe this to be the case. Hot topics or key discoveries in the external environment drive search within those same fields inside the organization (and drive hiring). However, our measure would not pick up more subtle processes of absorption that may involve combination of discoveries or know-how across sub-fields. So, for instance, let us assume that in the external environment, there is great deal of progress in sub-field A and sub-field B. Researchers inside the company see that by utilization advances from both of those sub-fields, they can make progress in sub-field C. This clearly represents absorption of know-how; however, our measure would not detect that as absorption. Given that prior research (e.g. Zander and Kogut 1995; Fleming, 2001) has identified combination of ideas from different fields as an important ingredient to innovation, this represents an important limit to our approach, and one that future research should address.

Our research, while shedding light on several factors that may influence absorption, also leaves many questions unanswered. First, there is clearly a large amount of unexplained variance. As mentioned above, a deeper exploration of the micro-level processes and management practices shaping absorption is clearly warranted by the significance of firm and laboratory level fixed effects. Second, for methodological reasons, we decided to leave out hotspot laboratories that came to the pharmaceutical firm purely through an acquisition. Our method depended on the

hotspot labs being greenfield in order to trace the evolution of each lab's know-how from a fixed point in time. However, it would be interesting in future research to examine the impact of corporate acquisition on the absorptive capacity of once-independent biotechnology companies located inside hotspots. The acquisition of smaller biotechnology firms by established pharmaceutical companies is a common strategy. How do such acquisitions impact absorption is a question that would have both practical significance as well as provide interesting theoretical insights on how changes in corporate control and governance impact innovative behavior.

Finally, this study was not designed to explore the overall performance implications of lab location strategies and absorption. We did not examine whether the laboratories with higher rates of absorption performed better (in terms of overall innovativeness) than laboratories with lower levels of absorption. Such a study involves a set of complex methodological challenges due to the very long time lags absorbing know-how the measurable manifestation of that know-how in the form of a drug that reaches the market (or even a later stage compound). As a result, the pharmaceutical industry may not actually be an ideal context to study the link between absorption and overall innovative performance. However, before distilling the normative implication of absorption, the field will need a much deeper understanding of the absorption phenomenon itself. To date, that understanding has been limited. We hope that the present study represents a helpful step in illuminating the phenomenon of absorptive capacity and its potential organizational and geographic drivers.

References

Adelman, D. E. and DeAngelis K. L. (2007) Patent Metrics: The Mismeasure of Innovation in the Biotech Patent Debate. Texas Law Review 85: 1677–1744.

Ahuja, G., & Katila, R. (2001) Technological acquisitions and the innovation performance of acquiring firms: A longitudinal study. Strategic Management Journal, 22: 197-220.

Almedia, P., and B. Kogut (1997) The exploration of technological diversity and the geographic localization of innovation. Small Business Economics 9:21-31

Andrews RL, Currim IS (2003) Recovering and profiling the true segmentation structure in markets: an empirical investigation. Int J Res Mark 20(2):177–192

Arabie P., Hubert L. (1994) Cluster analysis in marketing research. In: Bagozzi RP (ed) Advanced methods in marketing research. Blackwell, Cambridge, pp 160–189

Arora, A., Fosfuri, A. and Gambardella, A. (2001) Markets for Technology and Their Implications for Corporate Strategy. Industrial and Corporate Change, Oxford University Press, vol. 10(2), pages 419-51, June.

Audretsch, D. B. and Stephan P. E. (1999) Knowledge Spillovers in Biotechnology: Sources and Incentives. Journal of Evolutionary Economics 9 (1): 97-107

Baptista, R. (2000) Do innovations diffuse faster within geographical clusters?, International Journal of Industrial Organization, Elsevier, vol. 18(3), pages 515-535, April.

Baptista, R. and Swann P. (1998). Do firms in clusters innovate more? Research Policy. Elsevier, vol. 27(5), pages 525-540, September.

Belderbos, R., Carree, M., Diederen, B., Lokshin, B., Veugelers, R., (2004). Heterogeneity in R&D cooperation strategies. International Journal of Industrial Organization, 22 (8), 1237-1263, 455.

Best, M. (1990). The new competition: Institutions of industrial restructuring. Cambridge, UK: Polity.

Bishop, C.M. (2006) Pattern recognition and machine learning. Springer, Berlin

Breschi, S. & Lissoni, F. (2001). Knowledge spillovers and local innovation systems: A critical survey. Industrial and Corporate Change, 10(4), 975-1005.

Bryk, A.S., Raudenbush, S.W., (1992). Hierarchical Linear Models: Applications and Data Analysis Methods. Sage, NewburyPark, CA.

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, W. M. and Levinthal, D. A. (1989). Innovation and Learning - the 2 Faces of R-and-D. Economic Journal 99(397): 569-596.

Cohen, W. M. and Levinthal, D. A. (1990). Absorptive-Capacity - a New Perspective on Learning and Innovation. Administrative Science Quarterly 35(1): 128-152.

Cohen, W. M. and Levinthal, D. A. (1994). Fortune Favors the Prepared Firm. Management Science 40(2): 227-251.

Darr, E. P., Argote, L. and Epple, D. (1995). The Acquisition, transfer and depreciation of knowledge in service organizations: Productivity in franchises. Management Science, 41(11): 1750-1762.

Dyer, J. H. & Singh, H. (1998). The Relational View: Cooperative Strategy and Sources of Interorganizational Competitive Advantage. The Academy of Management Review, vol. 23, no. 4, pp. 660-679.

Feser, E. (1998). Enterprises, external economies, and economic development. Journal of Planning Literature 12:283-302.

Fleming, Lee. (2001). Recombinant Uncertainty in Technological Search. Management Science 47: 117-132

Galambos, L., and J. C. Sewell. (1995). Networks of Innovation. New York: Cambridge University Press.

Griliches, Z. (1979). Issues in assessing the contribution of research and development to productivity slowdown. Bell Journal of Economics 10:92-116.

Grossman, G. M., and E. Helpman. (1991). Innovation and growth in the global economy. Cambridge, MA: MIT Press.

J.J. Hox (2002). Multilevel Analysis. Techniques and Applications. Mahwah, NJ: Lawrence Erlbaum Associates.

Jaffe, A. (1986). Technological opportunity and spillovers of R&D: evidence from firm's patents, profits and market value. The American Economic Review 76:984–1001.

Jaffe, A. (1989). Real effects of academic research. American Economic Review 79:957-70.

Jaskowiak Pablo A., Campello R. J. G. B, Costa Filho, I. G. (2013). Proximity Measures for Clustering Gene Expression Microarray Data: A Validation Methodology and a Comparative Analysis. IEEE/ACM Transactions on Computational Biology and Bioinformatics.

Johnson, D. K. N. (2009) Not Far From the Madding Crowd: The Role of Proximity in Biotechnology Innovation. International Review of Business Research Papers 5(2): 420-429.

Kaufman L, Rousseeuw PJ (2005) Finding groups in data. An introduction to cluster analysis. Wiley, Hoboken, NY.

Krugman, P., (1991) Geography and Trade (MIT Press: Cambridge).

Lane, P., Koka, B. and Pathak, S. (2006) The reification of absorptive capacity: a critical review and rejuvenation of the construct. Academy of Management Review, 31, 4.

Lane, P. J. & Lubatkin, M. (1998). Relative Absorptive Capacity and Interorganizational Learning. Strategic Management Journal, vol. 19, no. pp. 461-477.

Liu, X., and White, S. (1997). The Relative Contributions of Foreign Technology and Domestic Inputs to Innovation in Chinese Manufacturing Industries, Technovation, Vol. 17, No. 3, pp. 119-125.

Marshall, A. (1920) Principles of Economics. London: Macmillan.

Maskell,P. (2001) Knowledge creation and diffusion in geographic clusters: Regional development implications. In: Felsenstein,D., McQuaid,R., McCann,D. and Shefer,D., (Eds.) Public Investment and Regional Economic Development pp. 59-76. Cheltenham: Edward Elgar.

Meeus et al., 2001 Meeus, M. T. H., Oerlemans, L. A. G., & Hage, J. (2001). Patterns of interactive learning in a high-tech region. Organization Studies, 22: 145–172.

Mowery, D. C. (ed.) (1988). International Collaborative Ventures in U.S. Manufacturing. Ballinger, Cambridge, MA.

Mowery, D. C., Oxley, J. E. and Silverman, B. S. (1996). Strategic alliances and interfirm knowledge transfer. Strategic Management Journal, 17, Winter special issue, 77–93.

Muscio, A. (2007). The impact of absorptive capacity on SMEs' collaboration. Economics of Innovation & New Technology, 16(8), 653-668.

Oltra, M.J. and M. Flor (2003). The Impact of Technological Opportunities and Innovative Capabilities on Firms' Output Innovation., Creativity & Innovation Management 12 (3), 137-145.

Owen-Smith, J., and W. W. Powell (2004). Knowledge networks in the Boston biotechnology community. Organization Science, 15: 5–21.

Pinch, S., N. Henry, M. Jenkins, S. Tallman (2003) From 'industrial districts' to 'knowledge clusters': a model of knowledge dissemination and competitive advantage in industrial agglomeration. Journal of Economic Geography, 3: 373-388.

Piore, M. J., and C. F. Sabel. (1984). The second industrial divide: Possibilities for prosperity. New York: Basic Books.

Pisano, G. P. (2006) Science Business: The Promise, the Reality, and the Future of Biotech. Boston: Harvard Business School Press.

Porter, M. (1990). The competitive advantage of nations. London: Macmillan.

Powell, W.W. (1998) Learning From Collaboration. Knowledge and Networks in the Biotechnology and Pharmaceutical Industries. California Management Review 40(3): 228-40.

Saxenian, A. (1994). Regional networks: Industrial adaptation in Silicon Valley and Route 128. Cambridge, MA: Harvard University Press.

Snijders, T. and Bosker, R.J. (1999). Multilevel Analysis: An introduction to basic and advanced multilevel modeling, London: Sage.

Stock, G.N., Greis, N.P., Fischer, W.A. (2001). Absorptive capacity and new product development. The Journal of High Technology Management Research 12, 77–91.

Szulanski G. (1996). Exploring internal stickiness: impediments to the transfer of best practice within the firm. Strategic Management Journal 17(1): 27–43.

Tsai, W. (2001) Knowledge Transfer in Intraorganizational Networks: Effects of Network Position and Absorptive Capacity on Business Unit Innovation and Performance. The Academy of Management Journal Vol. 44, No. 5, pp. 996-1004

USPTO Technology Assessment and Forecast Program, (1998) Technology Profile Report: Patent Examining Technology Center Groups 1630-1650, Biotechnology. United States Patent and Trademark Office report.

Van Den Bosch, F. a. J., Volberda, H. W. & De Boer, M. (1999). Coevolution of Firm Absorptive Capacity and Knowledge Environment: Organizational Forms and Combinative Capabilities. Organization Science, vol. 44, no. pp. 551-568.

Veugelers, R. (1997) Internal R&D expenditures and External Technology Sourcing, Research Policy, 26, 3, 303-316.

Zahra, S. & George, G. (2002). Absorptive Capacity: A Review, Reconceptualization, and Extension. Academy of Management Review, vol. 27, no. 2, pp. 185-203.

Zander, U., and Kogut, B. (1995) Knowledge and the speed of the transfer and imitation of organizational capabilities: An empirical test. Organization Science (6:1), pp. 76-92.

Zucker, L., Darby, M., and Armstrong, J. (1994) Intellectual Capital and the Firm: The Technology of Geographically Localized Knowledge Spillovers. Working Paper No. 9496. National Bureau of Economic Research, Cambridge MA.

Appendix

FIGURE 1. Lab's Technological Distances From Massachusetts in 2012

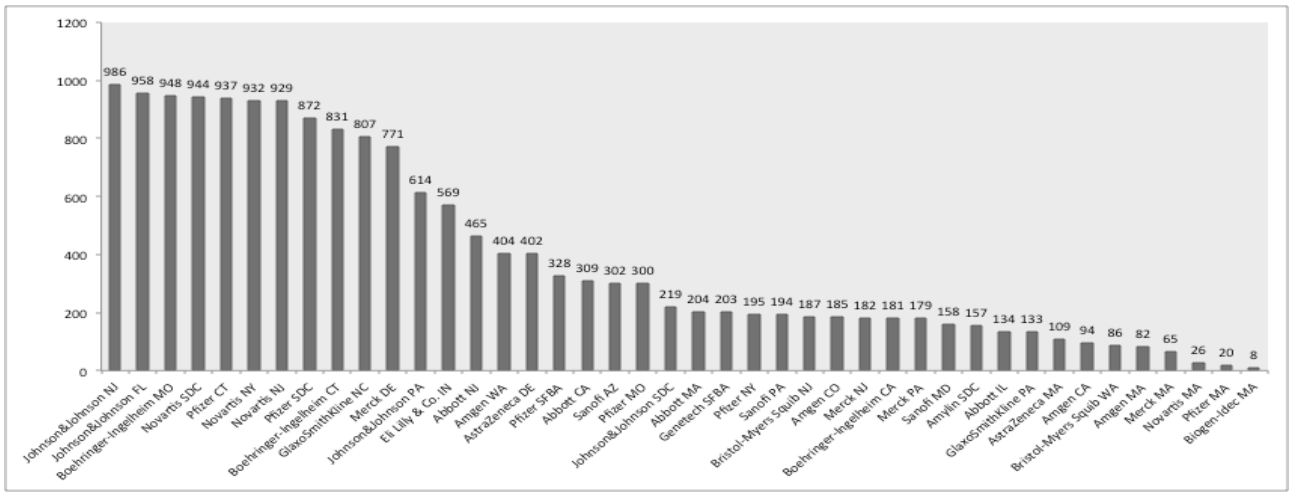

FIGURE 2. Lab's Technological Distances From San Diego County in 2012

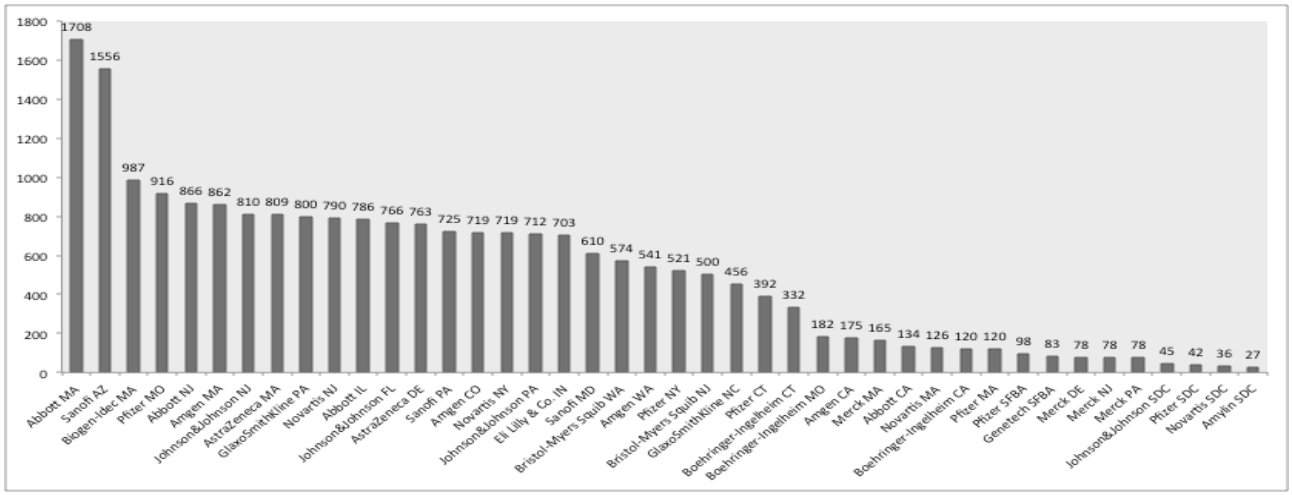

FIGURE 3. Lab's Technological Distances From San Francisco Bay Area in 2012

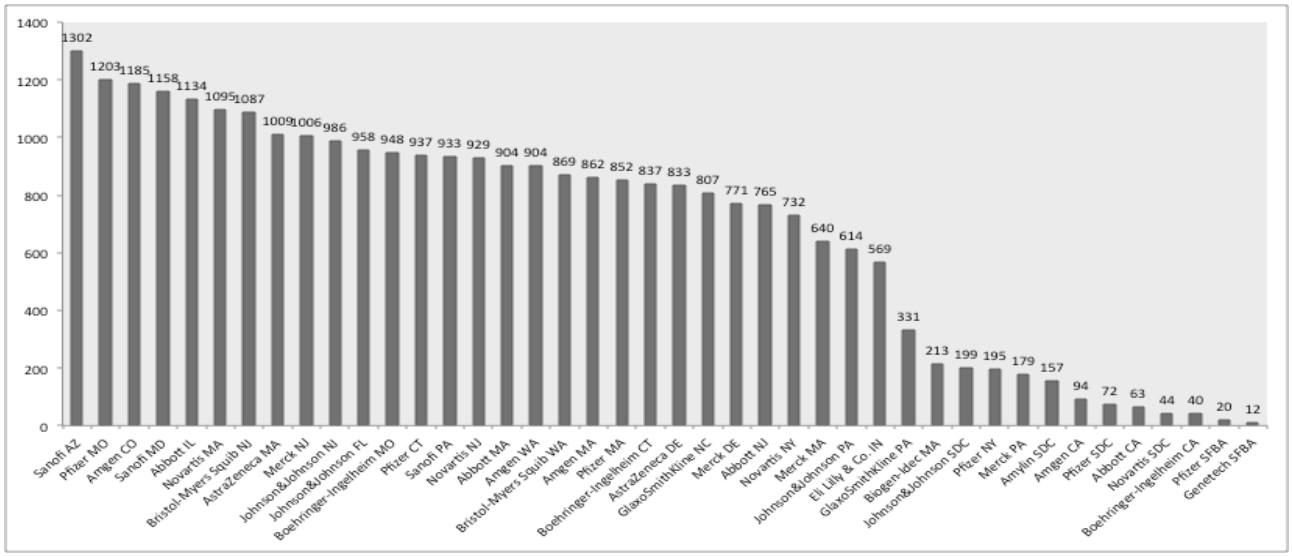

TABLE 1: R&D labs included in the analysis

Company	R&D Lab Location
Pfizer	MASSACHUSETTS SAN DIEGO COUNTY SAN FRANCISCO BAY AREA CONNECTICUT MISSOURI NEW JERSEY
Novartis	MASSACHUSETTS SAN DIEGO COUNTY NEW JERSEY NEW YORK
Merck	MASSACHUSETTS NEW JERSEY PENNSYLAVIA DELAWARE
AstraZeneca	MASSACHUSETTS DELAWARE
Johnson & Johnson	SAN DIEGO COUNTY PENNSYLVANIA NEW JERSEY FLORIDA
Abbott	MASSACHUSETTS ILLINOIS NEW JERSEY CALIFORNIA
Amgen	MASSACHUSETTS CALIFORNIA WASHINGTON COLORADO
Bristol-Myers Squibb	NEW JERSEY WASHINGTON
Eli Lilly	INDIANAPOLIS
Sanofi	MARYLAND ARIZONA PENNSYLAVIA
GlaxoSmithKline	PENNSYLVANIA NORTH CAROLINA
Boehringer-Ingelheim	CONNECTICUT MISSOURI CALIFORNIA

TABLE 2: Variables Description

Variable	Description
Technological Distance (Dependent)	Euclidean distance between a lab's technological fingerprint at time t and the target region's fingerprint at time t-1.
Log (Distance-Miles)	Log transformation of the geographical distance from the lab and the hotspot (in miles). - For the Massachusetts hotspot we have taken Kendal Square as geographical reference for the hotspot center-point - For the San Diego County hotspot we have taken the Scripps Research Institute as geographical reference for the hotspot center-point - For the San Francisco Bay Area we have taken the biotech company Genentech as geographical reference for the hotspot center-point
Distance Dummies	- Distance 1 Dummy variable equals to 1 if the distance between the lab and the hotspot is < or = to 5 miles; 0 if > than 5 miles - Distance 2 Dummy variable equals to 1 if a lab is located between 6 and 20 miles from the hotspot; 0 if > than 20 miles - Distance 3 Dummy variable equals to 1 if a lab is located between 20 and 50 miles from the hotspot; 0 if > than 50 miles - Distance 4 Dummy variable equals to 1 if a lab is located between 50 and 150 miles from the hotspot; 0 if > than 150 miles
Location Dummy	MA Dummy variable equals to 1 if the Lab is in MA; 0 if not
Time	Continuous variables indicating the specific year of observation
Lab Fixed Effects	Dummy variables equal to 1 if the observation belongs to Labi at time t; 0 if not
Firm Fixed Effect	Dummy variables equal to 1 if the Lab under observation belongs to Firmi at time t; 0 if not
Sister	Lagged variable indicating the average of the technological distance of other labs in the company network at the time t-1
Time*Log(Distance-Miles)	The interaction between time and distance

TABLE 3. Multilevel regression models of technological distance in Massachusetts

Parameter estimates:	Models Dependent Variable: Technological Distance to the MA Hotspot			
	I	II	III	IV
Constant	0.64* (0.11)	1.01* (0.63)	1.16* (0.22)	0.99* (0.12)
Log (Distance-Miles)	0.14* (0.18)		0.19* (0.20)	0.23* (0.16)
Distance Dummies - Distance 1 (= 1 if lab< 5 miles from hotspot) - Distance 2 (=1 is between 6 and 20 miles) - Distance 3 (= 1 if distance is between 20-50) - Distance 4 (= 1 if distance between 50-150 miles)		-0.46* (0.79) 0.97* (0.78) 0.27* (0.18) 0.88* (0.22)		
Location Dummy - MA			-0.45** (0.77)	-0.52** (0.98)
Time			-0.28** (0.10)	-0.33** (0.14)
Firm Fixed Effects			YES*	
Lab Fixed Effects			YES*	YES*
Patents			1.40 (0.44)	1.34 (0.36)
Time*Log_Distance-Miles			0.14* (0.02)	0.34* (0.01)
Sister Effects				-1.27* (0.10)
Model fit:				
deviance (-2LL)	2 915.30	2 543.41	2 774.73	2 544.42
Δdeviance (-2LL)	394.31	371.89	140.57	142.23
Δdf	9	1	1	1
Observations	915	915	915	915
Significance	**	**	**	**
pseudo-R^2	0.17	0.20	0.16	0.22
X^2				

** $p < 0.001$, * $p < 0.01$

TABLE 4. Multilevel regression models of technological distance in San Diego County

Parameter estimates:	Models Dependent Variable: Technological Distance to the SDC Hotspot			
	I	II	III	IV
Constant	0.44* (0.12)	1.02* (0.61)	1.25* (0.43)	0.89* (0.27)
Log (Distance-Miles)	0.24* (0.18)		0.98* (0.20)	0.23* (0.16)
Distance Dummies - Distance 1 (= 1 if lab< 5 miles from hotspot) - Distance 2 (=1 is between 6 and 20 miles) - Distance 3 (= 1 if distance is between 20-50) - Distance 4 (= 1 if distance between 50-150 miles)		 -0.46* (0.79) 0.98* (0.78) 0.27* (0.18) 0.88* (0.22)		
Location Dummy - MA			0.65** (0.78)	0.70** (0.81)
Time			-0.38** (0.18)	-0.47** (0.20)
Firm Fixed Effects			YES*	
Lab Fixed Effects			YES*	YES*
Patents			1.6 (0.69)	1.5 (0.84)
Time*Log_Distance-Miles			0.96* (0.12)	0.94* (0.16)
Sister Effects				-2.9* (1.84)
Model fit:				
deviance (-2LL)	2 815.32	2 443.41	2 674.73	2 444.43
Δdeviance (-2LL)	344.32	311.89	142.57	145.11
Δdf	9	1	1	1
Observations	915	915	915	915
Significance	**	**	**	**
pseudo-R^2	0.17	0.20	0.16	0.22

** $p < 0.001$, * $p < 0.01$

TABLE 5. Multilevel regression models of technological distance in San Francisco Bay Area

Parameter estimates:	Models Dependent Variable: Technological Distance to the SFBA Hotspot			
	I	II	III	IV
Constant	0.32* (0.11)	1.22* (0.21)	1.46* (0.23)	0.97* (0.17)
Log (Distance-Miles)	0.28* (0.37)		0.38* (0.45)	0.28* (0.12)
Distance Dummies - Distance 1 (= 1 if lab< 5 miles from hotspot) - Distance 2 (=1 is between 6 and 20 miles) - Distance 3 (= 1 if distance is between 20-50) - Distance 4 (= 1 if distance between 50-150 miles)		 -0.45* (0.77) -0.65* (0.43) -0.32* (0.18) 0.92* (0.32)		
Location Dummy - MA			0.55** (0.54)	0.62** (0.65)
Time			-0.18** (0.26)	-0.78** (0.57)
Firm Fixed Effects			YES*	
Lab Fixed Effects			YES*	YES*
Patents			1.9 (0.76)	1.7 (0.94)
Time*Log_Distance-Miles			0.92* (0.15)	0.82* (0.12)
Sister Effects				-0.49* (0.20)
Model fit:				
deviance (-2LL)	2 314.42	2 893.43	2 774.34	2984.42
Δdeviance (-2LL)	324.31	381.89	143.57	132.23
Δdf	9	1	1	1
Observations	915	915	915	915
Significance	**	**	**	**
pseudo-R^2	0.17	0.20	0.16	0.22

** $p < 0.001$, * $p < 0.01$