



Paper to be presented at the
DRUID Society Conference 2014, CBS, Copenhagen, June 16-18

Basic Science as a Prescription for Breakthrough Inventions in the Pharmaceutical Industry

Antonio Della Malva

KU LEuven

Managerial Economics and Strategy

antonio.dellamalva@kuleuven.be

Stijn Kelchtermans

KU Leuven (HUB)

Onderzoeksgroep globalisering, innovatie en competitie

stijn.kelchtermans@kuleuven.be

Bart Leten

KU Leuven

MSI

bart.leten@kuleuven.be

Reinhilde Veugelers

KU Leuven

Managerial Economics, Strategy and Innovation (MSI)

Reinhilde.Veugelers@kuleuven.be

Abstract

This analysis contributes to the understanding of the role of basic science in generating breakthrough inventions in the pharmaceutical industry. Recognizing the within-firm heterogeneity of inventive activities, we look not only at the firm level, but also at the firm-technology level for characteristics determining breakthroughs. A key finding is that firms pursuing basic science are more likely to produce breakthrough inventions. At the same time, doing more basic science in science disciplines that are closely linked to a given technology domain does not increase the likelihood of BTs in that particular technology. The insignificance of basic science intensity at the technology level, coupled to the significance at the firm level, suggests that the breakthrough rewards from science capacity are not reaped in the technology areas immediately involved in basic science, but in other areas of the technology portfolio of the firm. Our findings are

consistent with the view of science as a map to span processes of local search and the wider applicability of scientific insights.

1. Introduction

Since Schumpeter's seminal contributions (1939), economists have debated on the implications of different typologies of inventions. Inventions range from run-off-the-mill or incremental inventions to breakthrough or radical inventions. Whereas most of the measured productivity growth might come from incremental inventions (Baumol, 2004), breakthrough inventions (BTs), although more "rare events", introduce new concepts that have the potential to generate new markets and - through a mix of competitive and cooperative interactions - trigger follow-up inventions and growth in other firms. They are therefore critical building blocks of nations' creative destruction capacity and their long-term economic growth (Nelson & Winter, 1982; Malerba, 2004).

The debate on breakthroughs and their importance for corporate and country growth has remained rather rhetorical, only scantily backed up by empirical evidence. The lack of good empirical evidence is partly caused by a vague definition of the main concept and a lack of large scale empirical identification of breakthrough inventions moving beyond small numbers of case studies or self-reported measures (Henderson, 1993; Christensen, 1997; Schneider and Veugelers, 2010).

This paper contributes to a better understanding of the antecedents of breakthrough inventions, which are defined as those inventions with a high impact on subsequent inventive activity¹. The few empirical studies on breakthrough inventions have looked at the characteristics of the individual invention (e.g. Schoenmakers & Duysters, 2010; Arts &

¹As breakthrough inventions do not necessarily make it into commercial breakthroughs, our analysis cannot be extended into determinants of commercial breakthroughs. For an analysis of commercial breakthroughs in the Pharmaceutical industry, see a.o. Cockburn (2007).

Veugelers, 2012), firm characteristics (e.g. Ahuja and Lampert, 2001) or inventor characteristics (e.g. Singh and Fleming, 2010) most conducive to breakthroughs. Recognizing the within-firm heterogeneity of inventive activities, this study looks at the firm-technology level for characteristics determining breakthroughs. This research set-up allows a more comprehensive approach than most prior work by analysing simultaneously firm characteristics and firm-technology characteristics which are more prone to lead to breakthrough inventions.

In our quest for antecedents, we are particularly interested in the link to basic science as a predictor of high impact inventions. We examine science involvement both at the level of the firm and the level of the specific technology. This allows us to check “where” the benefits of basic science are realized. We do this for a large sample of firms in the pharmaceutical industry. We focus on this industry for two reasons. Firstly, innovation in the pharmaceutical industry is characterized by a high failure rate and large (mainly development) costs. This creates pressure to identify management practices that boost breakthrough diagnostics and therapeutics (Munos & Chin, 2011). Secondly, basic science is an important source for innovation in the pharmaceutical industry (Narin and Olivastro, 1992; Narin et al., 1997), with science insights resulting in the discovery of many breakthrough drugs (Cockburn and Henderson, 1998). The link between basic science and drug discovery has even increased over time, with companies moving from a mere “random screening” of compounds towards “informed” discovery methods fuelled by scientific knowledge (Pisano, 1997; Lim, 2004). While basic science is part of the drug discovery process of most firms, there is significant heterogeneity across firms in the basic science strategies adopted (Gambardella, 1995; Belderbos et al., 2010), providing room for further analysis of how basic science involvement may affect the breakthrough character of

inventive activities of firms. Taking into account that innovation in the pharmaceutical industry is carried out by dedicated biotech firms (DBFs) and large diversified pharmaceutical firms (LDPFs), who differ substantially in the way they tackle the innovation challenge, we also compare what determinants favor breakthrough inventions for each type of firm.

We characterize breakthroughs with varying intensities of impact distinguishing between inventions with an increasingly large footprint on subsequent inventive activity. To identify high-impact inventions, we look at the high end of the impact distribution, using benchmarks that are technology-specific i.e. our importance classification acknowledges that it is harder to 'stand out' in some technology domains than in others. Previously used measures tend to define high-impact inventions as those in a given top percentile or above a fixed cut-off value of the impact distribution without accounting for the variability of the entire distribution (Ahuja & Lampert, 2001; Singh & Fleming, 2010; Schoenmakers & Duysters, 2010; Kerr, 2010).

We find that firms pursuing basic science are more likely to generate breakthrough inventions. Interestingly, breakthrough inventions are not associated to higher scientific efforts targeted to the focal technology. The result – importance of scientific efforts at the firm level, but insignificance of targeted scientific efforts within the firm indicates that the breakthrough rewards of basic science are not reaped within technologies that are immediately involved in basic science, but in other technologies in a firm's portfolio. Our findings are consistent with a view of science as map (Fleming and Sorenson, 2004) which helps spanning the boundaries of a process of local search (Rosenkopf and Nerkar, 2001), and show that scientific insights often have practical value in many different technologies. With respect to other determinants of BTs,

we find that firms produce more breakthroughs in technologies in which they have disproportionately focused their technology portfolio and achieved technology leadership. Ceteris paribus, DBFs are more likely to generate breakthroughs than LDPFs.

The paper will first illustrate some previous findings from the literature on breakthrough inventions and pay special attention to two neglected dimensions: the within-firm heterogeneity across technologies and the role of basic science. We will present the data and detail the construction of the variables used in this study. The results will precede some concluding remarks.

2. Antecedents for Breakthrough Inventions

The most recent literature on breakthrough inventions defines them as foundational inventions that provide the framework for many subsequent inventions, contributing disproportionately to future technological developments (Rosenberg, 1994, Fleming 2001). Consistent with this definition, patent statistics can identify those inventions which are overly cited by later inventions. The empirical literature using highly cited patents to identify breakthroughs have mostly looked at the origins of these inventions. Although breakthroughs are often casted as revolutionary new ideas, they are often found to be much more evolutionary in origin, recombining already existing pieces of technological know-how.

New and small firms are considered to be at an advantageous position to generate breakthroughs. Older and larger firms, having to surmount an incumbency trap, are typically taken to be disadvantaged in introducing BTs. Breakthrough inventions render the capabilities of existing firms obsolete, putting them at a relative unfavourable position compared to follow-

up inventions on existing inventions in which they excel (Henderson, 1993; Shane, 2001). Empirical evidence from the radical innovation literature provides interesting insights on to the advantage of small firms in generating BTs (e.g. Schneider & Veugelers, 2010). But large firms are not necessarily doomed to incremental improvements of well-established technologies (Ahuja and Lampert, 2001; Chandy & Tellis, 2000), particularly if they can recombine existing competences with unfamiliar components into new applications (e.g. Fleming, 2001; Arts, 2012).

In one of the first quantitative firm-level studies on the topic, Ahuja and Lampert (2001) argue, among other things, that large firms can overcome innovative traps by experimenting with technologies which are either new to the firm or new to the technological domain. Fleming (2001) zooms in on this experimentation. He demonstrates that breakthroughs often originate from new re-combinations of the most familiar technological components, i.e. components which have been used extensively in the past but in different combinations. Schilling and Green (2011) analyse how different features of the search process underlying extreme impact ideas, namely search depth (the extensiveness of search within a particular domain), search scope (the breadth of a search across domains), and atypical connections, affect the likelihood of breakthrough idea generation. The authors stress the positive implications of deep reservoirs of knowledge in particular domains and the interaction of the latter with an extensive search breadth for impactful ideas to emerge. Agents should possess extensive knowledge of the domains in which critical solutions are sought and concurrently should extend the search to other domains. This finding is in line with a view of learning which pays off its highest dividend when it is directed beyond the actual technological and

organizational boundaries (Rosenkopf and Nerkar, 2001) and geared towards related familiar domains (Schilling et al., 2003). In line with this view, Phene et al. (2006) have highlighted the importance of related technical solutions in the case of biotechnology, noting however that too much exploration, spanning both technological and spatial boundaries at the same time, could turn harmful for the generation of breakthrough ideas.

Taken together, these findings from the literature suggest that the overall technology size, scope and specialisation pattern of firms will matter for generating breakthrough inventions. More particularly, they suggest within-firm heterogeneity across technologies with respect to the likelihood of breakthroughs. Firms might have different abilities to generate impactful inventions in technology domains, according to their relative strength in the diverse range of domains in which they operate and which they can leverage into other domains.

Basic Science and Breakthrough Inventions

According to the National Science Foundation (NSF) basic science is defined as the systematic study directed towards greater knowledge or understanding of the fundamental aspects of phenomena and observable facts without specific immediate commercial applications in mind (NSF, 2009). Corporate level basic science activities are concentrated in a small number of science-intensive industries, most prominently pharmaceuticals, electronics and aerospace (Mansfield, 1980; Rosenberg, 1990). Applied to the pharmaceutical industry, basic science includes attempts to reveal the mechanisms and processes of diseases, but does not include applied research activities such as compound screening, clinical trials and dosage testing (Lim, 2004).

Science, and basic science in particular, should matter for the generation of breakthrough inventions which most likely do not reside within the boundaries of the current paradigm and might entail a new recombination or shift in trajectory². Science entails a process of learning that generates knowledge on which firms can draw for their search into new territories.

Own science capabilities act as an admission ticket to R&D partnerships with university scientists and their networks (Liebeskind et al., 1996; Cockburn and Henderson, 1998; Murray, 2004). In order to plug in science networks, firms must have scientific competences to demonstrate their value as R&D partners (Hicks, 1994). Own scientific capabilities lead to a better identification, absorption and integration of external (public) knowledge (Cohen and Levinthal, 1990; Rosenberg, 1990; Gambardella, 1995; Cassiman and Veugelers, 2006). A firm's scientific capability serves as a map for technological landscapes guiding applied research in the direction of most promising technological avenues avoiding thereby wasteful experimentation (Rosenberg, 1990; Fleming and Sorenson, 2004). A better and more fundamental understanding of the technology landscape encourages non-local search for improving technologies as opposed to local search, leading to more diverse research trajectories being explored with higher potential pay-offs. Not only the average quality of the technologies is likely to be affected (Fabrizio, 2009): scientific capabilities might also affect the distribution of outcomes. Firms with scientific capabilities can be expected to generate more "unexpected" outcomes, leading to a higher probability of breakthrough inventions (Sobrero and Roberts, 2001).

² Throughout the text we refer to "Science" and "Basic Science" interchangeably as the arguments advanced in this work mostly relate to the process of developing knowledge to understand and explain natural phenomena, in contrast to "Applied Science". The latter is mostly geared towards the application of scientific principles to practical solutions (compounds in our case). In the context of Pharmaceuticals "applied science" is highly relevant too (e.g. clinical trials), but this type is not the main focus of attention in our analysis.

Although the literature offers a basic understanding of the process through which breakthroughs emerge and the role of science, no insights are provided on “where” breakthroughs are generated within the firm. Do breakthroughs happen in technology fields that are closely linked to the science fields where basic science is conducted, or do they manifest in other technologies in firms’ portfolios. Basic scientific research into disease mechanisms may lead to diagnostic or therapeutic insights with a broad scope of applications. More specifically, by looking beyond the immediate challenges posed by a given technology, basic science helps building a more comprehensive mental map of the problem at hand and may therefore be expected to contribute to fundamentally new approaches. This “reframing” function of basic science may lead to high-impact inventions *outside* the technology domains that have hitherto been linked to particular scientific disciplines. For example, while scientific advances in virology may traditionally have led to new vaccines (e.g. IPC class A61K), they have increasingly found applications in genetic engineering (e.g. IPC class C12N). In fact, the whole biotech (r)evolution is in itself a demonstration of this reframing principle. Based on these arguments, one could expect that the breakthrough rewards of basic science are realized within firms to a large extent in other technology fields than those that are immediately involved in the basic science.

3. Research design

3.1. Sample

To examine the relationship between basic science and breakthroughs, we use a panel of firms in the pharmaceutical industry. The sample firms are selected as the highest R&D spenders in the pharmaceutical industry based on the ‘2004 EU Industrial R&D Investment

Scoreboard'. This ranking lists the top 500 corporate investors in R&D whose parent is located in the EU, and the top 500 companies whose parent is located outside the EU (mainly US and Japan), based on corporate R&D expenditures in 2003. We match these firms with patent information from the European Patent Office (EPO). After excluding companies that specialize in medical instruments, we are left with 136 dedicated biotech firms (DBFs) and large diversified pharmaceutical firms (LDPFs) which applied for EPO patents between 1996 and 2002.³ This selection criterion retains in the sample the biotechnology firms who have played a key role in drug discovery in the past three decades (Danzon et al., 2005). The presence of these two types of firms enriches the data since, unlike the large Pharma firms, the small and entrepreneurial biotech firms often have their roots in public research and rely on markets for technology to commercialize their inventions rather than bringing new drugs to market themselves, except for the few large fully integrated biotech companies (FIBCOs).⁴

Patent data are used to measure the inventive activities of the sample firms. Patents are a good indicator of inventions in the pharmaceutical industry since the propensity to patent inventions is amongst the highest of all industries (Arundel and Kabla, 1998; Campbell, 2005). We rely on patent data from the European Patent Office (EPO). European patent data are preferred to the more commonly used data from the United States Patent and Trademark Office (USPTO). EPO patents are, typically, considered to provide a better indication of valuable inventions: the cost of patenting is two to five times higher at EPO; the workload of patent examiners is four times smaller at EPO; and the EPO has a 20-30% lower granting rate than the

³ The original database contains information on patenting activity of the sampled firms as of 1995. The first year is used to construct some of the control variables used in this study.

⁴ For a more elaborate discussion of the 'disintegration' trend in the pharmaceutical industry, see Cockburn (2007).

USPTO (Van Pottelsberghe de la Potterie & François, 2006; Quillen & Webster, 2001; Jaffe & Lerner, 2004). Since company names in patent databases are not harmonized and patents may be applied for under names of subsidiaries or divisions of a parent firm, we collected patent data at the consolidated parent firm level. Therefore, we searched, for each parent firm, for patents under the name of the parent firm as well as all their majority-owned subsidiaries. For this purpose, yearly lists of companies' subsidiaries included in corporate annual reports, yearly 10-K reports filed with the SEC in the US, and, for Japanese firms, information on foreign subsidiaries published by Toyo Keizai in the yearly 'Directories of Japanese Overseas Investments', were used. The consolidation was conducted on a yearly basis to take into account changes in the group structure of sample firms due to acquisitions, mergers, green-field investments and spin-offs. Acquisitions, and their patent stocks, are considered part of a parent firm from the year the acquisition transaction has been completed.

We restrict the analysis to those technologies where firms in our sample cover at least 1% of the patent applications at EPO. This restriction ensures that minor and unrelated technologies are excluded from the analysis. The 136 firms in the final sample account for 30,083 patent applications at the EPO in the observed 8-year period⁵. Our sample is composed of 77 DBFs and 59 LDPFs. 46% of our firm-technology-year observations are from DBFs - and almost 26% from firms with less than 250 employees. Most of the sample companies are based in the EU (61 out 136), 55 companies are US-based, 19 are from Japan and 1 firm is from Israel. Table A.1 reports the list of firms used in our sample, their geographical origins and the

⁵ The patent applications considered in the analysis have memberships in 108 3-digit IPC patent classes (out of 129 3-digit IPC classes). With the 1% coverage restriction, 83 technology classes are excluded.

typology. It includes both the largest among Pharma companies, such as Merck & Co. and Pfizer, next to large Biotech firms, such as Amgen, as well as small companies with few patents, most often in Biotech⁶.

3.2. Identifying Breakthrough Inventions

In line with earlier research, we use patent-based indicators for identifying inventions that are breakthroughs (e.g. Ahuja & Lampert, 2001; Singh & Fleming, 2010)⁷. Patents are widely used for invention analysis since they provide standardized information, suited for large scale analysis across time, country, technology and organisations. Patents are especially interesting for analysing breakthrough inventions and their antecedents since they leave a publicly accessible trail of backward and forward citations. These citations allow tracing, respectively, the inventions' origins and their technological impact on subsequent inventive activity.

Drawing the line between moderately influential inventions and 'extremely influential' or breakthrough inventions is not straightforward however. Breakthrough inventions have a disproportionate impact on future technological developments, but are rare events (Kuznets, 1962, as in Trajtenberg et al., 1997). Prior studies have identified breakthroughs as the top 1, 5 or 10 centiles of the forward citation distribution (Ahuja & Lampert, 2001; Singh & Fleming,

⁶ The share of generic firms is negligible, around 1% of the sample; they are not found to produce any breakthrough inventions, according to the definition of section 3.2. We therefore exclude this dimension from the analysis.

⁷ As an alternative to patent-citations, some scholars (e.g. Fontana et al., 2012) used data on R&D awards to identify breakthrough inventions. Comparing the number of patents received by award winning inventions and a control group, Carpenter et al. (1981) found that award winning inventions receive significantly more citations. This shows the close correspondence between citations and data on R&D awards to identify breakthrough inventions.

2010; Kerr, 2010) or those exceeding a fixed cut-off value of number of forward citations received (Schoenmakers and Duysters, 2010). A downside of these approaches is that they do not consider the variability of the entire forward citation distribution. For instance, in technologies where forward citations are evenly distributed across patents, the impact of “top centile inventions” may not be substantially higher than the impact of “moderately impactful” inventions.

In order to trace the inventions that truly stand out in a given technology domain in terms of impact, we take into account the dispersion of the forward citations around the mean for each 3-digit IPC patent class. For the purpose, we used the November 2011 version of the EPO/OECD Citation Database, which covers all patent applications (and their reference) at EPO since its foundation (Webb et al., 2005). More specifically, we calculate the number of forward citations received by the focal patent in the first 5 years after application from other patent applications at EPO, and define a breakthrough as a patent having a number of citations that is at least a number of standard deviations above the mean, for at least one 3-digit IPC patent class to which the patent belongs. As the reference scenario, we will be using 3 standard deviations, but to study the intensity of the breakthrough character of the invention, we will also look at 4 and 5 standard deviations for more extreme breakthroughs. While the standard deviation takes into account the dispersion of forward citations in the technology class, the mean controls for the importance of the technology class as such and for the propensity to patent in that particular technology. We believe that our approach defines a threshold for BTs which is less arbitrary than other definitions. The example below details the extent to which our approach differs from other approaches, such as the “top 1%” criterion”, where the threshold is

defined as the top 1st centile in the distribution of forward citations, in identifying BTs. For the most relevant technologies of the analysis, for instance, patent class A61 (MEDICAL OR VETERINARY SCIENCE; HYGIENE), C12 (BIOCHEMISTRY) and G01 (MEASURING, TESTING), the patents in the top 1% in each class receive less citations than our benchmark definition for BTs, mean plus 3 standard deviation. These are 3 out of 7 cases in which the top 1% of patent applications in a given patent class receive less citations than our benchmark definition. For the remaining 115 patent classes, our benchmark definition of BTs would be more inclusive than the “top 1%” criterion. Finally, for 34 patent classes, instead, our benchmark definition of BTs, is more inclusive than the defining a threshold at the 5th top centile of the distribution of forward citations.

Using patent citation information to measure impact restricts the analysis to assessing breakthroughs in the technological dimension (Henderson et al., 1998; Jaffe et al., 2000). As technological breakthroughs do not necessarily make it into commercial breakthroughs, our analysis cannot be extended into determinants of commercial breakthroughs⁸. Nevertheless previous research has indicated that forward citation patterns are related to social value (Trajtenberg, 1990) and private value (Harhoff et al, 1999; Hall et al., 2005).

3.3 Scientific Involvement of Firms

We look at measures to capture the involvement in science, both at firm-level and at firm-technology level. The simultaneous use of both measures allows us to examine whether the breakthrough rewards of science materialize in technologies that are directly linked to the scientific research, or in other technologies of the firm. We consider active involvement in

science through the number of scientific publications authored by researchers employed by the focal firm. To construct the science variables, we first extracted all publications in the Science Citation Index database of ISI/Thomson Scientific. Publication data is collected at the consolidated parent firm level, following a similar approach as the one followed for the collection of patent data. We used the CHI journal classification to identify publications on 'basic research'. The CHI journal classification assigns each of the SCI journals in one of four research levels, from very applied, targeted research to basic research. Biomedical journals are classified as 'clinical observation' (level 1), 'clinical mix' (level 2), 'clinical investigation' (level 3) and 'basic biomedical research' (level 4). Journals that are classified in level 4 are considered as reporting basic research findings. As we are interested in scientific capabilities, we will only consider the publications classified in level 4. Our firms file on average for 220 SCI publications per year. About 35% of the SCI publications of the sample firms are published in basic research journals and are used in the construction of the science variables.⁹

To calculate the specialization of scientific activities of our firms in technology classes, we matched the publications to 3-digit IPC technology classes through a concordance table between scientific disciplines¹⁰ and IPC classes that is created by Van Looy et al. (2004, 2006). The science-technology (ST) concordance is based on scientific references (almost 500'000) in granted EPO and USPTO patents applied between 1996 and 2001. Surveys of patent inventors (Tijssen, 2001; Fleming and Sorenson, 2004) have shown that patent inventors are aware of a significant share of the scientific papers cited in their patents, qualifying scientific non-patent

⁹ Our results are robust to adding a control for the applied science activities of firms (publications in levels 1-3).

¹⁰ Journal issues (and the articles they contain) are classified into one of 222 ISI Subject Categories. To increase the robustness of the concordance, the Expertise Centre for R&D Monitoring (ECOOM, KU Leuven) aggregated these detailed categories into 68 scientific disciplines.

references as indicators of the role of science in technology research, especially at an aggregated level such as technologies (Branstetter & Kwon, 2004; Fleming and Sorenson, 2004, Van Looy et al. 2007).

The ST concordance relates ISI science fields to 3-digit IPC technology classes via a set of probabilities. A high probability reflects that a science field is frequently cited in a particular technology domain. The probability linking science field *i* and technology field *j* is calculated as the ratio of “the number of scientific citations from patents in technology *j* to publications in science domain *i*” and the “total number of citations to publications in science field *i* across all technology fields”.¹¹ An example of a science field that is frequently cited in patents is *immunology*. This field receives most citations (>85 percent) from three IPC3 classes: A61 (*medical and veterinary technologies*), C07 (*organic chemistry*), C12 (*biochemistry*).¹²

Having identified basic research publications of the firms in our sample and having assigned them to technology classes to which they are mostly related, we measure the science intensity at the firm-level through the scientific publications in basic research authored by corporate scientists in the four years preceding the focal year per patent application in the previous four years (*Firm Science Intensity*). Publication variables are computed as four-year stocks to make the variables less prone to small number problems due to small numbers of annual publications for some sample firms. Looking at the publications as a ratio relative to the number of patents, eliminates the scale dimension when examining the science impact. To

¹¹ We have multiplied these probabilities by the overall technology intensiveness of science fields (i.e. the average number of times that publications in the field are cited in patents). We have also conducted analyses whereby we didn't apply this correction; this analysis gives very similar results.

¹² See Van Looy et al. (2004, 2006) for more details on the methodology to construct the ST concordance.

examine whether scientific capacity would matter not just at the aggregate firm level, but also at the specific focal technology level, we include the number of a firm's scientific publications related to the focal technology (using the concordance table, as explained above) relative to the number of firm patents in the technology between t-4 and t-1 (*Tech Science Intensity*). Whereas the first measure captures the overall involvement of the focal firm in basic science, the second measure provides an indication of how focused are the scientific capabilities of the focal firm to a specific technology area.

3.4 Firm and Firm-Technology Controls

We control for the technological size of the parent firm (*Firm Technology Size*), measured as the logarithm of the number of patent applications.¹³ It is usually highly correlated with firms' age, and aims at controlling for scale effects – i.e. scale economies, scope economies, etc. We also control for firms with less than 250 employees (*SME*).

To further characterize technology depth and breadth positions of firms, we need to determine how relevant is the inventive contribution of a particular firm for a technology and how central is the latter in the technological portfolio of a firm. We measure the leadership of a firm in a technology with respect to other competitors in the sample as the share of patent applications at the EPO by the focal firm in the focal technology class (*Tech Lead*). We also characterize the centrality which the focal technology has in the portfolio of the focal company as the number of patent applications in the focal technology divided by the total number of

¹³ A disadvantage of the use of patents as indicator of the scale of firms' inventive activities is that patents are an output indicator, which do not only reflect differences in inputs, but are also affected by patent propensities. However the propensity to patent is high in the pharmaceutical industry (Arundel and Kabla, 1998), making R&D expenditures and patent counts close substitutes for the scale of inventive activities in this particular industry.

patent applications by the focal firm (*Tech Specialization*). We also include a measure for the level of technology diversification in a firm's technology portfolio. Unlike the previous two measures, which are firm-technology specific, the latter is a firm-level one and synthesizes the extent to which firms marshal different technologies in their inventive process. Firms with a more diversified technology portfolio have access to a broader set of knowledge components to combine and create new inventions with potentially higher impact (Fleming, 2001). The variable is calculated as the Herfindahl index of the shares of the 3-digit IPC technology classes in the focal firm's patent portfolio (*Tech Concentration*). All the patent variables are 1-year lagged.

Finally, we control for differences related to the two main types of firms in the pharmaceutical industry. Dedicated biotech firms (DBFs) tend to differ systematically from the large diversified pharmaceutical (LDPFs) companies in the sense that they are created on the basis of their distinctive competencies in biotechnology, tend to have their roots in public research, rely on markets for technology rather than bringing new drugs to market themselves, and have fewer products in the market place to cannibalize (Orsenigo, 1989; Cockburn, 2007). We identify DBFs with a dummy variable (*Biotech*). We also control for whether the parent firm is headquartered in the United States of America (*US*), checking whether US firms have a competitive edge over non-US firms, being at home in the most important market for pharmaceuticals.¹⁴ Table 1 reports a description of the variables used in the study.

[INSERT TABLE 1 ABOUT HERE]

¹⁴ In 2011, North America accounted for 41.8% of world pharmaceutical sales compared with 26.8% for Europe (Efpia, 2012).

4. Descriptive Statistics

Table 2 reports the summary statistics on breakthroughs in our sample. Almost 5 out of 15 patents (the average size of the yearly technology-specific patent portfolio of a firm) do not receive any patent citation, being about 32% of the patents. About the same share of inventions receive more citations in the first five years from application at EPO than the average patent in the technology (*Above Average*). Breakthrough inventions are rare events, the more so, the more strict our definition for BTs. While on average 3% of the firms' yearly patents in a technology class are BT with the 3SD criterion, only 1% classify as BT with the 5SD criterion. .

[INSERT TABLE 2 ABOUT HERE]

The average number of breakthroughs is smaller in the biotech subsample than for the large pharmaceutical firms, which is not surprising since biotech firms are on average smaller in technology size, cf. *infra*. The absolute disadvantage of biotech becomes however smaller the more extreme the breakthroughs are. Moreover, in relative terms (breakthroughs as a share of overall patenting activity, also shown in Table 2), DBFs do not seem less apt to create breakthroughs than LDPFs, if not more apt.

[INSERT TABLE 3 ABOUT HERE]

[INSERT TABLE 4 ABOUT HERE]

Table 3 reports the summary statistics on our independent variables. On average, the firms in our sample have applied yearly for almost 52 patents (*Firm Tech Size*, exp. (3.949)) at

EPO between 1995 and 2002. The firms account on average for 0.3% of patent applications at the EPO in the technologies in which they operate (*Tech Lead*). Corporate scientists have published on average 1.4 basic science articles in the previous four years for every patent; despite producing in total less scientific articles than LDPFs, DBFs produce relatively more articles than patents: there are almost 3 articles for every 2 patent applications in the case of DBFs versus about 1 paper per patent in the case of LDPFs. *Tech science intensity* takes an average value of 1.6, and is significantly higher for LDPFs than DBFs.

The last column of Table 3 indicates that LDPFs dominate DBFs in terms of technology size and every covariate which is size dependent, which is why we scale other covariates where needed. Biotech firms are overrepresented among the SMEs and US firms in the sample. DBFs concentrate their inventive activities in a more pronounced manner relative to LDPFs: on average DBFs have a significantly higher share of their overall patenting activity in a specific technology class (*Tech Specialization*) than LDPFs. Their distribution over technology classes is however more balanced than it is for the LDPFs: *Tech Concentration* is significantly larger for the latter.¹⁵ Finally, Table 4 presents pair-wise correlations among the variables of our study. Expectedly, *Firm Tech Size* shows medium-to-high correlations with most other covariates. In addition, the measures of technological specialization (*Tech Specialization*) and scientific

¹⁵ Note that LDPFs exhibit a lower average value of *Tech specialization* than DBFs but greater *Tech concentration*. This stems from the fact that the technology class distribution for LDPFs is characterized by a relatively long right tail of technology classes with little activity, driving down average specialization (LDPFs are active in 8.4 technology classes on average versus 4.5 for DBFs). Combined with the relatively strong dominance of the A61 and C07 technology classes, which typically account for more than half of patents (see Table 5), this implies a high concentration for LDPFs.

specialization (*Tech Science Intensity*) are highly correlated, suggesting that firms concentrate their scientific activities in the neighbourhood of their core technological activities.¹⁶

Table 5 presents the ten most prolific firms in our sample with respect to generating absolute numbers of breakthroughs. It also adds five additional BT-prolific DBFs. Merck and Roche are LDPFs with the largest and most diverse technology portfolios, being active in more than 60 technology classes (*# IPC3*). Although they each account for about 5.7% of all BTs in the sample, these firms generate less BT compared to their technology size than Pfizer the most prolific in terms of absolute numbers of BTs generated in the sample. Astra Zeneca, not among the biggest LDPFs in terms of technology size and scope, is the most BT intense LDPF. Eli Lilly, a large LDPF, is not even in the top 10 for BTs.

[INSERT TABLE 5 ABOUT HERE]

DBFs are typically smaller in technology size and scope compared to LDPFs and as a consequence account for a lower absolute number of BTs. But they are typically more breakthrough-intense, i.e. hold a significantly higher share of breakthrough patents in their portfolio than LDPFs (*Share of Portfolio*).¹⁷ Incyte is the DBF which generates the largest absolute number of BTs, it also holds the largest technology size among DBFs.

All this shows the large heterogeneity in BT performance both between and within LDPFs and DBF. But Table 5 also shows an important heterogeneity within firms across technology areas, motivating our analysis at the firm-technology level: firms typically only

¹⁶ We thank the editor for having spotted this point.

¹⁷ Two-paired t-tests with unequal variance suggest that this difference is more striking the stricter is the criterion for breakthrough patents (the higher number of standard deviations to qualify for breakthroughs).

realize BTs in few of the technologies in which they are active (*# IPC3 with BT-3* versus *# IPC3*). This holds particularly for LDPFs, less for the DBFs. For example, Amgen and Millennium Pharma have the smallest scope among DBFs, but hold BTs in almost all sectors in which they are active. Finally, BTs as well as scientific publications (correlation between *Tech Specialization* and *Tech Science Intensity*) tend to be concentrated in the core technology domains of the firm (see the final two columns in Table 5).

5. Multivariate Analysis: Methods and Results

5.1 Methods

To explore the relationship between firms' activities in a technological domain and the generation of breakthrough inventions in the domain we use discrete choice models, specifically logistic regressions, as our dependent variable is the likelihood of introducing breakthrough inventions by a focal firm in a focal technology in a focal year. We are mainly interested in establishing partial correlations between the probability of generating a highly-cited patent in a given technology domain by a firm and a set of measures capturing the firms' scientific activities. BTs are concentrated in a handful of technology domains in which firms operate, rendering count data models difficult to estimate because of over-dispersion. As the most salient facet of the analysis lays in the likelihood of having BTs rather than in the number of breakthroughs, we prefer a logistic analysis rather than count data models. Given the nature of our data, which is organized in three dimensions – firm, technology and year –, we cluster standard errors at the technology-firm level to take into account dependence over time in the production of breakthrough inventions by firms in particular technologies. Finally, 25 technology dummies and 6 year dummies are included.

5.2 Results

The regression results for the likelihood of generating high-impact innovations are reported in Table 6. The results for BT-3 are our benchmark and basic science intensities at the firm and firm-technology level our major variables of interest.

We find that basic science involvement at the firm-level matters for generating breakthrough inventions (*Firm science intensity*). Firms with higher basic science intensity are more likely to produce breakthroughs: doubling firm's science intensity (at the mean) increases the likelihood of generating breakthroughs by about 18% in the case of BTs measured as 3 standard deviations above the mean. Furthermore, the magnitude of the effect raises when using more extreme definitions of BTs, – 22.5% for BT-4 and 23.4% for BT-5, both statistically different than the one for BT-3 at 10% significance level –, suggesting an increasing importance of science involvement on breakthroughs. At the same time, firms' involvement in basic science has no implications for the likelihood of having zero cited patents.

[INSERT TABLE 6 ABOUT HERE]

Yet, science involvement seems only to matter for breakthroughs at the firm level. Despite being concentrated in the core technologies of firms, where also BTs are mostly concentrated, technology-specific publications do not exert any impact on the likelihood of BT generation in that technology: *Technology Science Intensity* is not significantly associated with any measure of breakthroughs in the focal technology area. *Technology Science Intensity* is however significantly and negatively associated with the likelihood of generating inventions with no impact (zero forward citations), i.e. it helps to avoid zero impact inventions. The

insignificance of the technology specific science involvement for the likelihood of generating breakthroughs, coupled to the significance of the firm-level basic science variable is robust across alternative specifications.¹⁸ It suggests that the breakthrough rewards from science capacity are not reaped in the technology areas immediately implied in basic science, which is nevertheless the areas where most BTs occur, but in other areas of the technology portfolio of the firm. This is reminiscent of the role of basic science in supporting firms' search strategies for breakthroughs by recombining new pieces of information with existing familiar ones.

Technology size matters for breakthrough performance. Large technology firms (*Firm tech size*) are not only more likely to generate breakthroughs, but inventions of all kinds – from zero impact to the highest measured one. Once corrected for technology size, being an SME no longer matters, suggesting it is technology size, not employment size that matters for breakthrough performance. Given the technology size, the technology scope of firms (*Tech Concentration*), does not matter significantly for breakthrough performance.

To explain the within-firm heterogeneity in the likelihood of breakthrough generation across technologies, both *Tech Specialisation* and *Tech Lead* have a significant impact. Firms are more likely to generate breakthroughs in those technology areas where they hold a stronger technology lead over their competitors. Technology leaders in the field are also significantly more associated with zero cited patents. This is perhaps a reflection of a more defensive technology development strategy of incumbent leaders. At the same time, firms which are

¹⁸ We have also experimented with the inclusion of the share of publications in other technology fields than the focal one to the current specification and the substitution of Basic Science Intensity with the share of publications in other technology fields than the focal one. The results confirm the intuition that it is the involvement in science outside the focal technology which is associated with the generation of BTs. These additional results are available from the authors upon request.

specialized in the focal technology, i.e. for which the focal technology is core in the firm's overall technology portfolio, are more likely to generate breakthrough inventions in the focal technologies, as the coefficient associated to *Tech Specialization* is always positive and statistically significant across the spectrum of BT measures. Such firms are also more likely to generate zero cited inventions in the focal technology area. Once corrected for other firm and technology characteristics, US firms are significantly less likely to generate BTs but also zero-cited inventions. Last, but not least, although Biotech firms are generating less BTs in absolute terms, see section 4, the multivariate analysis, correcting for other firm and technology characteristics, shows that Biotech firms are disproportionately more likely to generate BTs as compared to LDPFs. This difference is sizeable and more pronounced for more extreme BTs.

We might expect the difference in breakthrough performance between biotech firms and traditional pharmaceutical firms to permeate beyond a fixed effect, as the firm and firm/technology characteristics might play out differently for the two types of firms. In Table 6 and Table 7 we show the results of the split sample by type of firm: LDPFs (Pharma) and DBFs (Biotech). The split results indicate that the firm-level science effect (*Firm Science Intensity*) remains positive across firms' typologies, but slightly larger for DBFs than for LDPFs. Indeed, doubling Firm Science Intensity at the mean results in a 16.27% increase in the likelihood of BT-3 for DBFs vis-a-vis a 14.92% increase for LDPFs. Instead, the effect of *Tech Science Intensity* on avoiding zero impact inventions reported in the total sample is attributable to LDPFs only. For DBFs there is no effect from technology-specific scientific activities.

[INSERT TABLE 7 ABOUT HERE]

[INSERT TABLE 8 ABOUT HERE]

What the split results also indicate is that DBFs which have a more concentrated technological portfolio generate more breakthroughs. The fact that this effect is only significant for DBFs might shed light on the different organization of the Biotech sector. Among LDPFs, having a concentrated technology portfolio does not allow to stand out in terms of breakthrough performance. It only helps to avoid failures as measured by zero-citation patents. Whereas biotech firms are generally active in fewer technology fields than LDPFs, some tend to concentrate more than others on a limited set of technology classes (cf. Amgen versus Incite or Genentech, see Table 3). What the multivariate analysis shows is that these specializing DBFs are, all else equal, more likely to generate BT inventions, compared to more diversified DBFs. But at the same time, and in contrast with LDPFs, these specializing DBFs are also more likely to generate zero impact inventions. These results would be consistent with specializing DBFs to be taken more risky research avenues. The split analysis also shows that the significant inferior breakthrough performance of US firms, all else equal, is mostly due to the LDPF subsample. In Biotech, US firms are not significantly underperforming with respect to breakthrough inventions. They are actually significantly better in avoiding zero impact inventions.

6. Conclusions

This analysis contributes to the understanding of the processes generating breakthrough inventions in the pharmaceutical industry. As opposed to previous studies, we take into account the within-firm heterogeneity of breakthrough performance by looking at the firm-technology level. We are particularly interested in the role of basic science in the generation of

breakthrough inventions. We examine science involvement both at the level of the firm and the level of specific technologies. This allows us to check “where” the benefits of basic science are realized. We operationalize breakthroughs with varying levels of impacts. We distinguish dedicated biotech firms (DBFs) and large diversified pharmaceutical firms (LDPFs) in our analyses, comparing which determinants favour breakthroughs for each type of firm.

We find that firms pursuing more intensively basic science, as measured by scientific publications authored by corporate researchers, produce more breakthrough inventions. Surprisingly, breakthrough inventions are not associated with higher scientific efforts targeted to the focal technology. This insignificance at the technology specific science involvement level, coupled to the significance of the firm level suggests that the breakthrough rewards from science capacity are not reaped in the technology areas immediately involved in basic science, but in other areas of the technology portfolio of the firm. This result is consistent with the ideas advanced by Nelson (1959) in his seminal paper on the *Economics of Basic Scientific Research*, where he stated that advances in basic scientific knowledge very often have practical value in multiple technology fields. Gibb’s law of phases, for example, inspired the design of equipment in technologies as diverse as petroleum refining, rubber vulcanization, nitrogen fixation and metal-ore separation (Nelson, 1959, p. 302).

Our findings inform managerial practice on basic science. Our study shows that the effects of basic science are not reaped in their own technology classes, but in other technology classes of the firm. This justifies assigning the responsibility for basic science at a central, corporate level. The role of central R&D departments is to complement the activities of

decentralized R&D units by sponsoring R&D projects that are of strategic importance for the company as a whole (Du et al., 2013).

Beyond science involvement, also firm's scale, scope and specialisation across technologies matter for breakthrough performance. In line with previous findings, the overall technology size of the firm matters for predicting overall inventive performance, including breakthrough performance. Beyond size, technology scope only matters to differentiate among DBFs in breakthrough performance. DBFs with more concentrated technology portfolios are more likely to generate breakthroughs. At the same time, they are also more likely to generate zero impact patents, consistent with a higher risk strategy. Firms are also more likely to generate breakthroughs in technology areas where they are specialized and hold a stronger technology lead over their competitors. After controlling for technology size and differences in basic science and technology strategies, DBFs generate more breakthroughs than LDPFs.

Our research has a number of limitations, which suggest directions for further research. In particular, one should be careful to generalize our findings to firms in other sectors.¹⁹ Besides the measurement problem that arises in industries with a lower propensity to patent, a specific concern is that knowledge creation and recombination in the pharmaceutical industry relies more on (basic) science than in other, less science-intensive industries (Narin et al., 1997; Narin and Olivastro, 1992; Rosenberg, 1990). Not only the mere availability of a large base of public, codified knowledge that inventors can draw on, but also the specific challenges in integrating this external knowledge with firm-specific knowledge may be issues that play out

¹⁹ Further note that our sample consists only of firms with (relatively) high R&D expenditures. Our results may therefore be less representative for the population of very small DBFs in the pharmaceutical industry.

differently in other industries. However, we suspect that our results do have relevance for firms in other industries in which science and technology development are closely linked, such as semiconductors and electronics (Pavitt, 1984; Klevorick et al, 1995). Consequently, it would be interesting to investigate whether the relations we have uncovered here hold up to scrutiny in non-science intensive industries. Further, we note as a caveat that we cannot fully rule out the possibility that the absence of a technology specific science effect might be partly caused by measurement errors in the science-technology concordance that is used for the creation of the basic science variables.

We plan to extend the current research, with a distinct interest to further pin down the role of scientific advancements for breakthrough inventions in drug discovery. Particularly the result of a firm level but no technology specific effect for science involvement discovered in the current analysis begs further investigation in which other than the focal technology areas firms can translate their scientific capabilities into breakthrough performance. This analysis would benefit from a more detailed level of technology analysis. Further research could also study which characteristics of basic science are amenable to breakthrough inventions. Do breakthrough inventions build only on radically new scientific insights, and how and how important are open, collaborative science modes to create breakthroughs? Finally, we reckon as a fruitful approach for further research the consideration of alternative conceptualizations of breakthroughs. More specifically, rather than using a simple count of a patent's forward citations, its impact may be captured by the breadth of its influence on future inventions, e.g. across technology domains. In other words, a technological invention may constitute a breakthrough if it leads the way to substantial further progress in either a specific technological

domain (“deepening”) or, as in the case of general purpose technologies, in a broad array of technologies (“widening”).

References

- Ahuja, G., & Morris Lampert, C. (2001). Entrepreneurship in the large corporation: A longitudinal study of how established firms create breakthrough inventions. *Strategic Management Journal*, 22(6-7), 521–543.
- Arts, S. (2012). Path Dependency and Recombinant Exploration: How Established Firms Can Outperform in the Creation of Technological breakthroughs. MSI working paper, 1228.
- Arts S., & Veugelers R. (2012). The technological origins and novelty of breakthrough inventions, MSI Mimeo
- Arundel A. and Kabla I. (1998). What percentage of innovations are patented? Empirical estimates from European firms. *Research Policy*, 27, 127-141.
- Baumol, W. (2004). Entrepreneurial enterprises, Large Established Firms, and Other Components of the Free-Market Growth Machine. *Small Business Economics*, 23(1): 9-21.
- Belderbos, R., Leten, B., & Kelchtermans, S. (2010). Performing and Exploiting Basic Research in the World's Major Pharmaceutical Firms. MSI Research Paper 1003, KU Leuven.
- Branstetter, L., Kwon, H., (2004). The restructuring of Japanese research and development: The increasing impact of science on Japanese R&D. RIETI Discussion Paper Series 04-E-021.
- Campbell J.J. (2005). Understanding Pharma. A Primer on How Pharmaceutical Companies Really Work. Pharmaceutical Institute, Inc., Raleigh, NC.
- Carpenter, M.P., Narin, F., & Woolf, P. (1981). Citation rates to technologically important patents. *World Patent Information*, 3, 160-163.

- Cassiman, B., & Veugelers, R. (2006). In Search of Complementarity in Innovation Strategy: Internal R&D and External Knowledge Acquisition. *Management Science*, 52(1), 68–82.
- Chandy, R. K., & Tellis, G. J. (2000). The Incumbent's Curse? Incumbency, Size, and Radical Product Innovation. *Journal of Marketing*, 64(3), 1–17.
- Christensen, C. (1997). *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail*. Harvard Business Press.
- Cockburn, I., & Henderson R., (1998). Absorptive capacity, coauthoring behavior, and the organization of research in drug discovery. *The Journal of Industrial Economics* 46(2), 157-182
- Cockburn, I. (2007). Is the Pharmaceutical Industry in a Productivity Crisis? In: *Innovation Policy and the Economy*, Volume 7 (Lerner, J., Stern, S., Eds.). MIT Press, p1-32.
- Cohen, W. M., & Levinthal, D. A. (1990). Absorptive Capacity: A New Perspective on Learning and Innovation. *Administrative Science Quarterly*, 35(1), 128–152.
- Danzon, P. M., Nicholson, S., & Pereira, N. S. (2005). Productivity in pharmaceutical–biotechnology R&D: the role of experience and alliances. *Journal of Health Economics*, 24(2), 317–339.
- Du, J., Leten, B. & Vanhaverbeke, W. (2013). *Managing Open Innovation with Science-Based and Market-Based Partners*. Mimeo.
- Efpia, (2012). *The Pharmaceutical Industry in Figures. Key Data 2012*. European Federation of Pharmaceutical Industries and Associations.
- Fleming, L. (2001). Recombinant Uncertainty in Technological Search. *Management Science*, 47(1), 117–132.

- Fleming, L., & Sorenson, O. (2004). Science as a map in technological search. *Strategic Management Journal*, 25(8-9), 909–928.
- Fabrizio, K.R. (2009). Absorptive capacity and the search for innovation. *Research Policy*, 38(2), 255-267.
- Fontana, R., Nuvolari, A., Shimizu, H., & Vezzuli, A. (2012). Schumpeterian patterns of innovation and the sources of breakthrough invention: evidence from a data set of R&D awards. *Journal of Evolutionary Economics*, 22, 785-810.
- Gambardella, A. (1995). *Science and Innovation: The US Pharmaceutical Industry During the 1980s*. Cambridge University Press.
- Hall, B. H., Jaffe, A., & Trajtenberg, M. (2005). Market Value and Patent Citations. *The RAND Journal of Economics*, 36(1), 16–38.
- 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. (1993). Underinvestment and Incompetence as Responses to Radical Innovation: Evidence from the Photolithographic Alignment Equipment Industry. *The RAND Journal of Economics*, 24(2), 248–270.
- Henderson, R., Jaffe, A. B., & Trajtenberg, M. (1998). Universities as a Source of Commercial Technology: A Detailed Analysis of University Patenting, 1965–1988. *Review of Economics and Statistics*, 80(1), 119–127..
- Hicks, D., Ishizuka, T., Keen, P., Sweet, S., (1994). Japanese corporations, scientific research and globalization. *Research Policy* 23, 375-384.

- Jaffe, A. B., Trajtenberg, M., & Fogarty, M. S. (2000). The Meaning of Patent Citations: Report on the NBER/Case-Western Reserve Survey of Patentees. National Bureau of Economic Research Working Paper Series, No. 7631.
- Jaffe, A., Lerner, J. 2004. Innovation and Its Discontents: How Our Broken Patent System Is Endangering Innovation and Progress, and What to Do about It. Princeton, NJ: Princeton University Press.
- Kerr, W. R. (2010). Breakthrough inventions and migrating clusters of innovation. *Journal of Urban Economics*, 67(1), 46–60
- Klevorick, A., Levin, R., Nelson, R., Winter, S. (1995). On the sources and significance of inter-industry differences in technological opportunities. *Research Policy*, 24, 185–205.
- Liebeskind, J.P., Oliver, A.L., Zucker, L., Brewer, M., (1996). Social networks, learning, and flexibility: sourcing scientific knowledge in new biotechnology firms. *Organization Science* 7(4), 428-442.
- Lim, K., (2004). The relationship between research and innovation in the semiconductor and pharmaceutical industries (1981-1997). *Research Policy* 33, 287-321.
- Malerba, F. (2004). *Sectoral Systems of Innovation*. Cambridge University Press, Cambridge, MA.
- Mansfield, E., 1980. Basic research and productivity increase in manufacturing. *The American Economic Review* 70(5), 863-873.
- McMillan S., Narin F., & Deeds D. (2000). An Analysis of the Critical Role of Public Science in Innovation: the case of Biotechnology. *Research Policy*, 29, 1-8.

- Munos, B. H., & Chin, W. W. (2011). How to Revive Breakthrough Innovation in the Pharmaceutical Industry. *Science Translational Medicine*, 3(89)..
- Murray, F., (2004). The role of academic inventors in entrepreneurial firms: Sharing the laboratory life. *Research Policy* 33(4), 643-659.
- Nelson, R., (1959). The simple economics of basic scientific research. *The Journal of Political Economy* 67(3), 297-306.
- Nelson, R. R. & Winter, S. (1982). *An evolutionary theory of economic change*. Cambridge: Harvard U. Press.
- Narin, F. & Olivastro, D. (1992). Search report – linkage between technology and science. *Research Policy* 21(3), 237-249.
- Narin, F., Hamilton, K., Olivastro, D. (1997). The increasing linkage between U.S. technology and public science. *Research Policy* 26, 317-330.
- National Science Foundation, (2009). *Business R&D and Innovation Survey*.
- Orsenigo, L. (1989). *The emergence of biotechnology: institutions and markets in industrial innovation*. Pinter, 230p.
- Pavitt K. (1984). Sectoral patterns of technical change – Towards a taxonomy and theory. *Research Policy*, 13(6), 343-373.
- Phene A., Fladmoe-Lindquist K. & Marsh L. (2006). Breakthrough Innovations in the US Biotechnology Industry: the effects of technological space and geographic origin. *Strategic Management Journal*, 27, 369-388.
- Pisano, G., (1997). *The development factory*. Boston (MA): Harvard Business School Press.

- Quillen, C., Webster, O., (2001). Continuing Patent Applications and Performance of the US Patent Office. *Federal Circuit Bar Journal* 11, 1–21.
- Rosenberg N. (1990). Why do firms do basic research (with their own money)? *Research Policy*, 19 (2), 165–174
- Rosenberg N. (1994). *Exploring the Black Box: Technology, Economics, and History*. Cambridge University Press.
- Rosenkopf, L., Nerkar, A. (2001). Beyond local search: boundary-spanning, exploration, and impact in the optical disk industry. *Strategic Management Journal*, 22(4), 287–306.
- Schilling M., Green E. (2011). Recombinant search and breakthrough idea generation: An analysis of high impact papers in the social sciences. *Research Policy*, 40 (10), 1321–1331.
- Schilling M., Vidal P., Ployhart R.E. & Marangoni A. (2003). Learning by doing something else: Variation, Relatedness and the Learning Curve, *Management Science*, 49(1), 39-56.
- Schneider, C., & Veugelers, R. (2010). On Young Highly Innovative Companies: Why They Matter and How (not) to Policy Support Them. *Industrial and Corporate Change*, 19(4), 969–1007.
- Schoenmakers, W., & Duysters, G. (2010). The technological origins of radical inventions. *Research Policy*, 39(8), 1051–1059.
- Schumpeter, J. (1939). *Business Cycles. A Theoretical, Historical and Statistical Analysis of the Capitalist Process*. 2 vols. New York: McGraw-Hill.
- Shane, S. (2001). Technology Regimes and New Firm Formation. *Management Science*, 47(9), 1173–1190.

- Singh, J., & Fleming, L. (2010). Lone Inventors as Sources of Breakthroughs: Myth or Reality? *Management Science*, 56(1), 41–56.
- Sobrero, M., & Roberts, E. B. (2001). The Trade-off between Efficiency and Learning in Interorganizational Relationships for Product Development. *Management Science*, 47(4), 493–511.
- Tijssen, R., (2001). Global and domestic utilization of industrial relevant science: patent citation analysis of science-technology interactions and knowledge flows. *Research Policy* 30, 35-54.
- Trajtenberg, M. (1990). A Penny for Your Quotes: Patent Citations and the Value of Innovations. *The RAND Journal of Economics*, 21(1), 172–187.
- Trajtenberg, M., Henderson, R., & Jaffe, A. (1997). University Versus Corporate Patents: A Window On The Basicness Of Invention. *Economics of Innovation and New Technology*, 5(1), 19–50.
- Van Looy, B., Tijssen, R.J.W. Callaert, J, Van Leeuwen, T., Debackere, K., 2004. European science in industrial relevant research areas: Development of an indicator-based bibliometric methodology for performance analyses of countries and research organizations, Report for the European Commission (DG Research) produced by the Centre for Science and Technology Studies, Leiden, (CWTS) and International Centre for Studies in Entrepreneurship and Innovation Management, Leuven (INCENTIM).
- Van Looy, B., Debackere, K., Callaert, J., Tijssen, R., Van Leeuwen T., 2006. Scientific capabilities and technological performance of National Innovation Systems: An exploration of emerging industrial relevant research domains. *Scientometrics* 66(2), 295-310.

- Van Looy, B., Magerman, T., Debackere, K., (2007). Developing technology in the vicinity of science: An examination of the relationship between science intensity (of patents) and technological productivity within the field of biotechnology. *Scientometrics* 70(2), 441-458.
- Van Pottelsberghe de la Potterie, B., François, D., 2006. The Cost Factor in Patent Systems. EPO working paper CEB 06-002.
- Webb, C., Harhoff, D., Hoisl, K., 2005. Analyzing European and international patent citations: a set of EPO database building blocks. OECD Science, Technology and Industry Working Paper No. 2005/9, Paris.

Tables

Table 1: Variable Description

VARIABLE	DESCRIPTION	LEVEL*
Zero	Number of patents by the focal firm in the focal technology class that have NOT received any citations in the first 5 years after application at EPO	FTY
BT 3	... that have received more citations than THREE standard deviation above the average number of citations in the first 5 years after application at EPO	FTY
BT 4	... that have received more citations than FOUR standard deviation above the average number of citations in the first 5 years after application at EPO	FTY
BT 5	... that have received more citations than FIVE standard deviation above the average number of citations in the first 5 years after application at EPO	FTY
Firm Tech Size	Logarithm of the number of patent applications at EPO by the focal firm in year t-1	FY
Tech Lead	Share of patents in the focal technology by the focal firm at EPO in year t-1	FTY
Tech Specialization	Share of patents in the focal technology in the technological portfolio of focal firm in year t-1	FTY
Tech Concentration	Herfindahl Index of shares of technology classes in the technological portfolio of the focal firm calculated between t-4 and t-1	FY
Firm Science Intensity	Number of scientific articles published by focal firm's scientists between t-4 and t-1 divided by the number of patent applications between t-4 and t-1	FY
Tech Science Intensity	Number of publications relevant for the technology published by the focal firm's scientists between t-4 and t-1 divided by the number of patent applications in the technology between t-4 and t-1	FTY
US	Dummy variable for US firms	F
Biotech	Dummy variable for Dedicated Biotech Firms (DBFs)	F
SME	Dummy variables for firms with less than 250 employees	FY

F indicates that the variable is at the firm-level; FY indicates that the variable is at firm-year level; FTY indicates that the variable is firm-technology-year.

Table 2: Summary statistics for breakthroughs

Number of patents per firm/year/technology class	All firms (3,080 obs)			LDPF (1,648 obs)		DBF (1,432 obs)	
	mean	% of total	sd	mean	% of total	mean	% of total
with 5-year fwd citations = 0	4.77	32%	11.31	6.81	33%	2.42	30%
with 5-year fwd citations > mean + 3 sd ("BT 3")	0.41	3%	1.23	0.49	2%	0.3	4%
with 5-year fwd citations > mean + 4 sd ("BT 4")	0.23	2%	0.79	0.27	1%	0.19	2%
with 5-year fwd citations > mean + 5 sd ("BT 5")	0.13	1%	0.51	0.14	1%	0.12	1%
Total	14.71	100%	33.72	20.41	100%	8.14	100%

T-tests on the differences of the means between LDPF and DBF are all statistically significant at the 1% level.

Table 3: Summary statistics for firm/technology characteristics

	All Sample		LDPFs		DBFs		Difference ¹
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	
Firm-Technology²							
Tech Lead	0.003	0.007	0.004	0.008	0.002	0.005	+ ***
Tech Specialization	0.222	0.187	0.192	0.205	0.256	0.157	- ***
Tech Science Intensity	1.609	3.367	1.870	4.085	1.312	2.251	+ ***
Firm³							
Firm Tech Size	3.949	1.571	4.672	1.488	3.123	1.217	+ ***
Tech Concentration	0.328	0.128	0.332	0.132	0.323	0.122	+ ***
Firm Science Intensity	1.374	2.287	1.196	4.085	1.578	3.081	- ***
US	0.449	0.497	0.330	0.470	0.583	0.493	- ***
SME	0.216	0.441	0.066	0.248	0.489	0.500	***

¹ t-tests on the differences of the means of variables between LDPF and DBF. *** indicates statistical significance at the 1% level.

² Statistics for firm-technology level variables have been calculated from 3,080 observations (1648 for LDPFs and 1432 DBFs)

³ Statistics for firm-level variables have been calculated from 136 observations (59 LDPFs and 80 DBFs)

Table 4: Correlation Table

	BT-3	Firm Tech Size	Tech Lead	Tech Specialization	Tech Concentration	Basic Science Intensity	Tech Publication Share	US	Biotech	SME
BT-3	1									
Firm Tech Size	0.325	1								
Tech Lead	0.370	0.427	1							
Tech Specialization	0.132	-0.508	-0.015	1						
Tech Concentration	-0.057	-0.334	-0.151	0.416	1					
Firm Science Intensity	0.057	-0.130	-0.081	0.086	-0.011	1				
Tech Science Intensity	0.022	-0.565	-0.069	0.718	0.409	0.068	1			
US	0.061	0.127	0.082	-0.043	-0.048	0.015	-0.049	1		
Biotech	-0.087	-0.489	-0.206	0.192	-0.047	0.103	0.212	0.292	1	
SME	-0.145	-0.488	-0.183	0.258	0.171	-0.015	0.289	0.016	0.431	1

Table 5: Distribution of Breakthroughs: Top 10 Firms & Largest Dedicated-Biotech Firms

PARENT	Type	Patents	BT-3	Share of Portfolio	# IPC3	# IPC3 with BT-3	Top-IPC3 with most BT3 **	Share of Top-IPC3 in Patent Portfolio	Share of BT-3s in Top-IPC3
PFIZER	LDPF	1619	50	3.09%	34	10	A61	86%	94%
ASTRAZENECA	LDPF	770	37	4.81%	24	6	A61	84%	100%
MERCK CO	LDPF	2429	36	1.48%	63	17	A61	66%	69%
ROCHE	LDPF	2207	36	1.63%	61	19	A61	57%	58%
GLAXOSMITHKLINE	LDPF	1126	30	2.66%	39	13	A61	84%	80%
INCYTE CORPORATION	DBF	555	29	5.23%	13	7	G01	62%	86%
BRISTOL MYERS SQUIBB	LDPF	931	28	3.01%	37	6	A61	86%	86%
GENENTECH INC	DBF	449	19	4.23%	14	6	A61	82%	26%
HUMAN GENOME SCIENCES	DBF	565	19	3.36%	10	8	A61	88%	63%
WYETH	LDPF	917	18	1.96%	20	5	A61	85%	78%
VERTEX PHARMACEUTICALS	DBF	177	14	7.91%	10	4	A61	81%	93%
MILLENNIUM PHARMACEUTICALS	DBF	399	13	3.26%	8	6	A61	66%	62%
CHIRON	DBF	393	12	3.05%	16	6	A61	80%	75%
AMGEN	DBF	251	7	2.79%	7	5	A61	91%	57%
ELAN	DBF	198	7	3.54%	16	4	A61	91%	71%

Variable Description. *Type*: LDPF stands for Large Diversified Pharmaceutical Firm and DBF for Dedicated Biotech Firm; *Patents*: number of patent applications between 1995 and 2002; *BT-3*: Number of breakthrough patents with forward citations beyond three standard deviations from the technology average; *Share of Portfolio*: Share of BT-3 in firms' patent portfolio; *#-IPC3*: Number of IPC-3 patent classes in which the focal firm applied for patents; *# IPC3 with BT-3*: Number of IPC3 in which the focal firm has breakthrough patents; *Top-IPC3 with most BT3*: IPC-3 where the focal firm has the majority of breakthroughs; *Share of Top-IPC3 in Patent Portfolio*: Number of patents in the IPC-3 with most BT-3 as share of all firm patents; *Share of BT-3s in Top-IPC3*: Number of BT-3 in the IPC-3 with most BT-3, as share of all BT-3's of the firm.

**A61: Medical or Veterinary Science; C12: Biochemistry; Beer; Spirits; Wine; Vinegar; Microbiology; Enzymology; Mutation Or Genetic Engineering; G01: Measuring; Testing

Table 6: Logit Estimations (All Firms: 3080 observations, 695 firm-technology clusters)

	Dependent variable = Likelihood of patents with 5-year forward citations in one of the following categories			
	0	> mean + 3 sd	> mean + 4 sd	> mean + 5 sd
Firm Tech Size	0.922*** (0.083)	1.053*** (0.088)	1.050*** (0.096)	1.194*** (0.114)
Tech Lead	129.033** (64.703)	86.883*** (21.33)	105.586*** (22.609)	57.590*** (17.085)
Tech Specialization	2.326*** (0.562)	2.757*** (0.577)	2.128*** (0.598)	2.805*** (0.739)
Tech Concentration	0.039 (0.509)	0.894 (0.681)	1.409* (0.777)	1.332 (1.034)
Firm Science Intensity	0.017 (0.019)	0.166*** (0.039)	0.193*** (0.046)	0.172*** (0.051)
Tech Science Intensity	-0.059*** (0.021)	-0.021 (0.022)	-0.041 (0.036)	-0.028 (0.04)
US	-0.386*** (0.130)	-0.366** (0.143)	-0.295* (0.155)	-0.4** (0.168)
Biotech	0.013 (0.159)	0.692*** (0.183)	1.047*** (0.201)	1.227*** (0.237)
SME	-0.239* (0.140)	-0.11 (0.197)	-0.247 (0.233)	-0.416 (0.305)
Intercept	-3.529*** (0.87)	-8.868*** (0.874)	-9.262*** (0.937)	-10.385*** (1.078)
YEAR DUMMIES	YES	YES	YES	YES
TECHNOLOGY CLASS DUMMIES	YES	YES	YES	YES
Pseudo R ²	0.219	0.285	0.308	0.299

Table 7: Logit Estimations (Pharmaceutical Firms: 1648 observations, 355 firm-technology clusters)

	Dependent variable = Likelihood of patents with 5-year forward citations in one of the following categories			
	0	> mean + 3 sd	> mean + 4 sd	> mean + 5 sd
Firm Tech Size	0.979*** (0.130)	1.077*** (0.120)	1.093*** (0.122)	1.229*** (0.147)
Tech Lead	86.152 (63.374)	75.709*** (26.10)	86.887*** (25.833)	32.363 (20.319)
Tech Specialization	2.049* (1.117)	1.713 (1.070)	2.628** (1.043)	3.424*** (1.31)
Tech Concentration	-2.418* (0.995)	0.873 (1.02)	1.412 (1.233)	0.281 (1.559)
Firm Science Intensity	-0.095 (0.097)	0.210*** (0.069)	0.242*** (0.086)	0.187** (0.089)
Tech Science Intensity	-0.067* (0.035)	-0.042 (0.04)	-0.083 (0.058)	-0.057 (0.063)
US	-0.065 (0.251)	-0.652*** (0.193)	-0.518** (0.203)	-0.471** (0.207)
SME	-0.001 (0.35)	-0.707 (0.446)	-0.201 (0.533)	-0.021 (0.667)
Intercept	-3.285*** (0.975)	-8.552*** (1.016)	-8.996*** (1.047)	-10.029*** (1.243)
YEAR DUMMIES	YES	YES	YES	YES
TECHNOLOGY CLASS DUMMIES	YES	YES	YES	YES
Pseudo R ²	0.249	0.325	0.328	0.301

Table 8: Logit Estimations (Biotechnology Firms: 1432 observations, 336 firm-technology clusters)

	Dependent variable = Likelihood of patents with 5-year forward citations in one of the following categories			
	0	> mean + 3 sd	> mean + 4 sd	> mean + 5 sd
Firm Tech Size	0.915*** (0.097)	1.242*** (0.145)	1.107*** (0.16)	1.37*** (0.208)
Tech Lead	271.847*** (90.346)	104.3031*** (30.521)	155.387*** (34.082)	107.372*** (34.529)
Tech Specialization	1.951*** (0.655)	1.6794** (0.757)	0.127 (0.866)	0.739 (1.120)
Tech Concentration	1.822*** (0.702)	2.4524*** (0.888)	1.884* (1.052)	3.467*** (1.285)
Firm Science Intensity	0.045 (0.029)	0.1307*** (0.049)	0.159*** (0.058)	0.185** (0.08)
Tech Science Intensity	-0.015 (0.037)	0.0333 (0.041)	0.024 (0.061)	0.051 (0.080)
US	-0.476*** (0.163)	0.0065 (0.265)	0.012 (0.310)	-0.3938 (0.388)
SME	-0.2387 (0.158)	0.2019 (0.239)	-0.098 (0.267)	-0.351 (0.353)
Intercept	-3.1816*** (0.465)	-8.2682*** (0.731)	-8.191*** (0.821)	-9.1284*** (0.944)
YEAR DUMMIES	YES	YES	YES	YES
TECHNOLOGY CLASS DUMMIES	YES	YES	YES	YES
Pseudo R ²	0.189	0.278	0.313	0.331