# In search of anti-commons: patent-paper pairs in biotechnology. An analysis of citation flows.

**Tom Magerman**[a,b,c]

**Bart Van Looy**[a,b,c]

**Koenraad Debackere**[a,b,c]

[a] *Managerial Economics, Strategy and Innovation, Faculty of Business and Economics, K.U.Leuven, Leuven (Belgium)*
[b] *Research Division INCENTIM, K.U.Leuven, Leuven (Belgium)*
[c] *ECOOM, Leuven (Belgium)*

**Abstract**

In this paper, we examine the possible presence of anti-commons dynamics in biotechnology. Patent-paper pairs - i.e. scientific publications from which the contents (methodology, findings, discovery) is part of a patent application - have been detected by relying on text mining algorithms. Starting from a dataset consisting of 1,025,005 scientific publications and 119,016 EPO and USPTO patent documents, a total of 584 patent-paper pairs have been identified. In terms of scientific citations, publications with a patent counterpart receive considerable more publication from other publication than publications without a patent counterpart. This is not the case for the technological impact of patent-paper pairs; forward patent citation rates do not differ significantly between patents with or without a scientific counterpart. As such our findings do not provide evidence for the presence of anti-commons effects stemming from the introduction of IP within scientific activities in the field of biotechnology.

# Introduction: Entrepreneurial Universities

Collaboration between science and industry, and the phenomenon of 'enterprising universities', have been studied extensively over the last few decades. This growing interest is connected to the increasing acknowledgement of the fundamental role of knowledge and innovation in stimulating technological performance, international competitiveness and economic growth. Researchers in the domain of innovation (including Freeman, 1987 and 1994; Lundvall, 1992; Nelson, 1993; Nelson and Rosenberg, 1993; Mansfield and Lee, 1996; Mansfield, 1995; Mowery and Nelson, 1999; Dosi, 2000) stress the role of science and the importance of interaction between a variety of institutional actors underlying the innovative capacity and consequent economic performance of an economical system. This more encompassing view on innovation dynamics has resulted in a growing popularity of the 'innovation system' concept which gained acceptance by scholars and policy makers alike as a guiding framework to understand innovation dynamics on an aggregated level (OECD, 1999; European Innovation Scoreboard, 2002).

In these models, knowledge generating institutions such as universities, research laboratories, industrial research centres and more recently government institutions are acknowledged – besides firms and entrepreneurs - as important players in developing and stimulating the innovative capacity of a particular region or country. Likewise, the Triple Helix model, which emerged in the second half of the 1990s (Leydesdorff and Etzkowitz, 1996 and 1998; Etzkowitz and Leydesdorff, 1997), emphasizes both the complementary roles of firms, knowledge creation institutes - including universities - and governmental agencies, as well as the importance of constructive interactions among them.

There are multiple reasons why universities are relevant actors within innovation systems and can contribute to the national innovative capacity. First, research institutions produce information and ideas upon which the development of new products, processes and services can build. Secondly, research institutions can work on certain research agendas for a longer period of time, which can lead to the creation of new scientific insights. The latter can over time lead to economic applications. Notice in this respect that universities are well placed to address market failures that occur in the field of innovation (Arrow, 1962; Freeman, 1994; Baumol, 2002). Such market failures arise especially in relation to basic research, characterized not only by high levels of uncertainty both in terms of technical and commercial success, but also spanning long time frames to bear fruit (often decades). In addition, the nature of the outcomes of innovative activity - i.e. knowledge or information - complicates investment decisions even further (Foray, 2004). All these phenomena pose specific challenges for private investors, who tend to refrain from becoming involved in basic research activities. In order to avoid a loss of social welfare – due to non-investment behaviour of private firms – most national innovation systems nowadays invest considerably in basic research performed at universities and public research institutes.

As such, knowledge institutions like universities can play a specific role related directly to the potential these institutions possess to avoid technological lock-in phenomena. In order to continuously stimulate economic growth within a particular region or nation, based on knowledge intensive entrepreneurship, its technology portfolio should strike a balance between routine technological activities on the one hand (these are focused on process and incremental development in the more mature phases of the technology life cycle) and non-routine technological activities on the other hand (these are more focused on new technology platforms and fundamental developments). Local / regional knowledge centres, especially universities and research centres, can play a significant part in this respect. As they participate in high level scientific research, they contribute to the generation of new knowledge. Such research takes place in international research communities. The exploration of new fields of knowledge[1] – that can often not yet be categorized as routine activities – and the continued diffusion of this knowledge among regional actors can be considered an essential task of knowledge centres and especially universities. This double dynamic allows knowledge centres to play a fundamental role in regional innovation networks. These institutions are best placed to offer support in regard to the dual challenge of local and global knowledge development (Debackere, 2000; Van Looy, Debackere and Andries, 2003; Lester and Piore, 2004; Debackere and Veugelers, 2005). If a particular region fails to include this dual task as a priority in their regional innovation policy, there is a long term risk of regression and growth stagnation due to the life cycle phenomenon. It is in this context that

[1] Innovative economic activities imply a process of cross-fertilization in which different knowledge domains are involved. Knowledge centers with a large variety of disciplines consequently have greater potential for cross-fertilization. By further developing this potential, they can greatly contribute to preventing the risks of technological mono-cultures.

the significance of knowledge centres should be seen: they also develop non-routine activities in research communities which participate in knowledge exchange on an international scale. As such, universities offer regions exploration possibilities that are essential for mid to long term innovation potential. Lester points in this respect to the importance for innovation of 'interpretative', problem defining activities, besides analytical, problem solving ones. When enterprises focus on the latter, it is essential that sufficient attention is paid to creating an environment for exploration. In this sense, universities, as fora where new ideas can be explored and studied, are indispensable.

These reflections also imply that universities are more effective in this respect as they are more active in scientific research. Recent research in the US as well as in Europe confirms this relation: an explicit research focus coincides with a larger number of enterprising activities (patents, spin-offs, contract research) (di Gregorio and Shane, 2003; O'Shea and Allen, 2006; Van Looy et al., 2005; Sapsalis et al., 2006).

At the same time, contributing effectively to the innovative capacity of an innovation system requires a willingness of universities to become more 'entrepreneurial'. The notion of 'entrepreneurial universities' (Branscomb, Kodama and Florida, 1999; Etzkowitz, Webster and Healy, 1998) refers to the development of the following spectrum of activities: more intense commercialization of research results, patent and license activities, spin-off activities, collaboration projects with the industry, and greater involvement in economic and social development. As such, one observes a 'second academic revolution'[2] whereby education and research become complemented with service and valorisation activities aimed at transferring new scientific knowledge to economical activity realms.

Indeed, nowadays an increasing activity of academic researchers in exploiting their discoveries can be observed (Henderson, Jaffe and Trajtenberg, 1998; Thursby and Thursby, 2002; Meyer, Sinilainen and Utecht, 2003; Lissoni et al., 2008) and university patents become an important – and visible - method of technology transfer (Basberg, 1987; Schwartz, 1988; Boitani and Ciciotti, 1990; Trajtenberg, 1990; Archibugi, 1992).

Interaction and exchange between academia and industry can result in positive aspects, both for the business partner (e.g. Zucker and Darby, 2001; Hall, Link and Scott, 2001; Faems,

[2] During the first academic revolution (19th century) research became a part of universities activity profile.

Van Looy and Debackere, 2005) and for the academic sector (e.g. realization of complementarities between applied and basic research – Azoulay, Ding and Stuart, 2009; Callaert et al., 2009; generation of new research ideas – Rosenberg, 1998; attracting additional resources for (basic) research - Agrawal and Henderson, 2002). Additional benefits – when introducing IP in scientific activities - can be found in the facilitation of the creation of a market for ideas and the ability of society to realize the commercial and social benefits of a given discovery (Kitch, 1977; Merges and Nelson, 1990; Gans and Stern, 2000; Arora, Fosfuri and Gambardella, 2004; Hellman, 2007; Murray and Scott, 2007).

At the same time some concerns arise due to the increasing commercialization of scientific activities undertaken by universities. First, too much emphasis on (market) exploitation might negatively impact the quantity and quality of scientific research. While a complete crowding out of scientific activities by commercialization endeavours is considered as highly unlikely (Merton, 1968; Scotchmer, 2004; Thursby, Thursby and Gupta-Mukherjee, 2007), some scholars however do signal a (moderate) negative impact on the quality of research (Henderson, Jaffe and Trajtenberg, 1996; Trajtenberg, Henderson and Jaffe, 1997; Czarnitzki, Glänzel, and Hussinger, 2009). At the same time, a majority of reported empirical findings report a positive relationship between patenting and publication outcomes of academic researchers (Fabrizio and DiMinin, 2008; Van Looy et al., 2006, Breschi, Lissoni and Montobbio, 2007; Czarnitzki, Glänzel and Hussinger, 2007; Stephan et al., 2007). Patents as commercialized discoveries seem to be rather by-products of scientific work than substitutes (Murray, 2006).

While most empirical evidence – on the level of individual scientists – reports a positive relationship between patenting activities and publication outcomes (quantity as well as quality), the expansion of IPR might still result in 'privatizing' the scientific commons and potentially limiting scientific progress (Argyres and Liebeskind, 1998; David, 2000; Krimsky, 2004). This fear is nicely expressed by the metaphor of the "Tragedy the anti-commons", introduced by Heller (Heller, 1998) as opposed to the "Tragedy of the commons" of Hardin (Hardin, 1968). Heller states that the presence of too many owners with blocking power can lead to the underutilization of scarce resources, or, translated to the world of IPR, more intellectual property rights may lead paradoxically to fewer useful products (too many owners hold rights in previous discoveries creating obstacles for future research). Although this phenomenon is induced by high transaction costs and can be transitional (market players have to learn to deal with each other or changing market circumstances), patent anti-commons could prove more intractable in biomedical research than in other settings because of the importance of patents for the biotechnology industries, the lack of substitutes for certain biomedical discoveries (rivals may not be able to invent around) and the heterogeneity of interests and resources among public and private patent owners (Heller and Eisenberg, 1998).

To the extent such anti-commons effect exists, one can however wonder whether IPR and the exploitation of scientific research in se is the problem, or the enforcement in specific circumstances and the behaviour of licensors (Walsh, Arora and Cohen, 2003; Murray, 2006). A recent policy forum article by Van Overwalle (2010) further illustrates how this "anti-commons effect" can indeed be dealt with through the design of appropriate exemption or exclusion policies coupled to the design of patent pools.

Although anecdotal evidence exists of problematic use of IPR on scientific findings (e.g. the 'OncoMouse' or 'Havard mouse' of Leder and Stewart; patents on human genes associated with breast and ovarian cancer owned by Myriad Genetics), large scale evidence of the presence of an anti-commons effect in biotechnology patenting is rare. One notable exception is the study of Murray and Stern (2007) suggesting a modest anti-commons effect based on a decline in citation rate – after granting of the patent - by 10 to 20% for a set of 169 patent-paper pairs published in Nature Biotech between 1997 and 1999, although these authors also clearly point to the interpretation limits inherent to their study.

In our study, we want to contribute to the research on an anti-commons effect in biotechnology by comparing citation patterns of patents and scientific publications for a large dataset containing all biotechnology patents (EPO and USPTO) and scientific publications (published in ISI Web of Science covered journals) from 1991 to 2008. First we investigate whether biotechnology publications for which a counterpart exists in the patent system (so called 'patent-paper pairs', scientific publications from which the contents - methodology, findings, discovery - is part of a patent application) are cited differently (more/less) within scientific journals, compared to similar biotechnology publications which are not related to a patent document. Next, we engage in a similar analysis focusing this time on 'technological' citations: to what extent are patents closely related to scientific publications cited differently by other patents compared to biotechnology patents without scientific counterpart. The former will allow us to shed some light on the fear that exploitation of scientific findings is hampering scientific development by pruning promising developments due to the introduction of (potentially blocking) patents. The latter will allow us to look at the technological impact of scientific developments that become translated into a patent.

An important methodological aspect for this kind of studies relates to the identification of those patent-paper pairs, scientific publications for which a patent equivalent is present. To obtain a broad set of patent-paper pairs, we stepped down from a manually guided process of mapping patent and scientific publications and developed a new approach of automated, large scale, mapping of patents and scientific publications based on content similarity by relying on text mining algorithms. This approach allows large scale processing of patents and scientific publications to detect patent-paper pairs.

Within the next pages, we first outline the selection of the data used for this analysis, followed by a description of the methodology adopted to assess the similarity between patents and scientific publications. This section is followed by reporting the findings, for scientific citations and patent citations respectively. We conclude with outlining the limitations of our work and suggest avenues for further research in this area.

# Data and methodology

## *Field selection*

We focus on patents and scientific publications in the field of biotechnology because it is a field well known for the presence of science-technology linkages and because the large scale exploitation of biomedical research makes it more susceptible to an anti-commons effect (Heller and Eisenberg, 1998).

Patents and publications are selected based on technological and scientific classification schemes respectively. Patent-paper pairs are identified by matching the content of patent and scientific publication's title and abstract by using text mining algorithms.

## *Selection of biotechnology patents*

On the patent side, the OECD definition of biotechnology is used to identify biotechnology patents (OECD, 2005), defining 30 International Patent Classification subclasses/groups related to biotechnology (see Appendix A for the list of IPC-subclasses/groups used for the selection). We use PATSTAT (EPO Worldwide Patent Statistical Database) to retrieve all EPO and USPTO granted patents with application and grant year between 1991 and 2008 according to the 30 defined IPC-subclasses/groups related to biotechnology. This led to a set of 27,241 EPO and 91,775 USPTO patents (PATSTAT edition October 2009). As text mining techniques are applied for the further identification of patent-paper pairs, only patents with titles and a minimum abstract length of 250 characters were withheld, resulting in a final patent data set of 7,254 EPO and 80,994 USPTO biotechnology patents (hence 88,248 patents in total).

### *Selection of scientific publications*

On the publication side, we select biotechnology publications (articles, letters, notes, reviews)[3] from the Thomson Reuters ISI Web Of Science database based on the Web of Science subject classification, for the same time period 1991-2008 (volume year). 243,361 publications are revealed from subject category 'Biotechnology and Applied Microbiology'. However, to ensure that all potentially related scientific publications are present in the data set, we extend this 'core' publication set with publications from nine related subject categories: 'Biochemical Research Methods'; 'Biochemistry & Molecular Biology'; 'Biophysics'; 'Plant sciences'; 'Cell Biology'; 'Developmental Biology'; 'Food sciences & Technology'; 'Genetics & Heredity' and 'Microbiology Materials'[4]. This results in more than 1.75 million additional publications for the period 1991-2008 - a considerable computational challenge for the text mining method to identify patent-paper pairs. To lower the number of publications for ease of calculations without losing too much relevant documents, we only retain those publications from this extended set that are citing or are cited by at least one publication from our core set, sizing down the extended publication set to 683,674 publications. Finally we also add all multidisciplinary publications from 'Nature', 'Science' and 'Proceedings of the National Academy of Sciences of the United States of America', resulting in 97,970 additional publications. Again we only retain publication documents with titles and a minimum abstract length of 250 characters, resulting in a final publication set of 948,432 biotechnology related publications.

---

[3] Articles are by far the biggest category (90% articles compared to 1.5% letters, 2% notes and 6.5% reviews)

[4] The authors want to thank Wolfgang Glänzel for his kind help in the development of a search strategy for biotechnology publications.

## *Text mining oriented identification of patent-paper pairs*

The identification of patent-paper pairs is based on the content similarity of titles and abstracts of patents and publications, derived by text mining algorithms (Latent Semantic Analysis - Landauer et al., 2007). For all patents, the similarity with all publications is derived based on content similarity metrics. Patent-paper combinations with similarity scores beyond validated thresholds are retained as patent-paper pairs under the condition that at least one of the patent inventors is listed as publication author.

In practice, first all titles and abstracts of all biotechnology patents and scientific publications where indexed[5]. During indexing, a limited number of stop words (123) were removed, stemming was applied (Porter stemmer) and all terms only occurring once in the corpus were removed. Next, a vector space model (Salton, Wong and Yang, 1975) was created based on the full text index[6]. This vector space consists of a document by term matrix, whereby rows are defined by all included documents and columns consists of all (stemmed) terms identified within the set of documents. Overall, the matrix used in this analysis consists of 1,066,632 rows (documents) and 301,697 columns (terms). Within a next step, multiple similarity metrics were derived, based on different variants of weighting (e.g. TF-IDF), different levels of data (or dimensionality) reduction (SVD) and finally similarity measure (for more details on those options, see Magerman et al., 2009). For this study, three different options have been pursued: a binary weighting scheme, a scheme based on Inverted Document Frequencies and a scheme based on the combination of Term Frequencies and Inverted Document Frequencies. In terms of data or dimensionality reduction, 9 variants based on Singular Value Decomposition were used – ranging from 5 to 1,000 dimensions to retain. For the similarity measure, a classic cosine measure was used as well as a simple count based on the number of

---

[5] Apache-Lucene™, an open source text search engine library, was used for the indexing (http://lucene.apache.org/java/docs/index.html)

[6] MatWorks Matlab™, a commercial packet for technical computing, was used for the construction of the vector space and further data handling and similarity calculation (http://www.mathworks.com/products/matlab/). The authors want to thank Frizo Janssens who was so kind to share his propriety Matlab code for the import of the full text index into a document-by-term matrix.

common terms. The combination of these options resulted in a set of 43 measures[7], which have been used to calculate the similarity between all documents within the dataset (patents and publications).

A thorough manual validation of 300 cases was set up to select the metric best suited for the identification of patent-paper pairs. This validation effort revealed that dimensionality reduction as advocated by methods like Latent Semantic Analysis underperforms compared to a normal cosine measure on the full data, and measures based on a mere count of the number of common terms yields the best and most robust results in terms of identifying patent-paper pairs without missing relevant pairs.

Two metrics were combined for the classification of patent-paper combinations. The number of common terms, divided by the minimum of the number of terms of the patent document on the one hand and of the publication document on the other hand, is used for a first selection of patent-paper combinations with significant content similarity (CommonTermsMin $\geq$ 0.60). A second criterion, based on the number of common terms divided by the maximum of the number of terms of the patent document and publication document, is used to filter out ambiguous cases (CommonTermsMax $\geq$ 0.30). These two content-based criteria are combined with an additional criterion based on ownership: at least one of the patent inventors has to be listed as a publication author. Together those three criteria allow an accurate identification of patent-paper pairs (precision equals to 0.96/0.98 and recall equals to 0.90/0.84 – depending on a conservative or optimistic definition of equality - based on the data of the validation sample of 300 cases). For a more elaborate discussion of the technical details and comparison of content based similarity measures best suited for the identification of patent-paper pairs, we refer to Magerman et al., 2011.

### *Identified patent-paper pairs*

The starting point for the identification of patent-paper pairs is the combined dataset of 88,248 biotech patents and 948,432 biotech publications[8]. Application of the first matching

[7] 9 levels of dimensionality reduction plus one variant without dimensionality reduction are combined with 4 weighting methods, resulting in 10x4=40 measures. Three variants based on the number of common terms are added (three variants of normalization), resulting in a total of 43 measures.

criterion, a content similarity of at least 0.60 based on the number of common terms weighted for the minimum of the number of terms of both documents, yields 27,250 related patent-paper combinations out of the more than 80 billion combinations under examination. Application of the second matching criterion, a content similarity of at least 0.30 based on the number of common terms weighted for the maximum of the number of terms of both documents, resulted in 645 patent-paper pairs. Application of the last criterion, at least one patent inventor being listed as a publication author, resulted in a final set of 584 patent-paper pairs. 17 patents are matched with multiple publications (up to three publications), which seems to be cases of (partly) disclosure of the same results in multiple scientific articles. At the same time, 115 publications are matched to multiple patents (up to seven patents), which revealed to be members of the same patent family. Hence we have 566 distinct biotechnology patents having a paired biotechnology publication, and 400 distinct biotechnology publications having a paired biotechnology patent.

Note that we deliberately opted for a very conservative selection to identify patent-paper pairs. Especially the second criterion filters out a lot of ambiguous cases, so we can be confident that the described patent-paper matching method reveals real patent-paper combinations.

## Findings on citation patterns of scientific publications (publication-to-publication citations)

Within this section we report and discuss the empirical results obtained when analysing scientific citations - i.e. citations from other scientific publications - to scientific publications that are part of a patent-paper pair. This analysis implies a comparison with scientific citations to scientific publications which do not belong to a patent-paper pair.

### *Descriptive statistics*

Table 1 provides a summary overview of the number of biotechnology publications under study as well as the observed forward citations, organized by publication year.

---

[8] Only patents and publications with titles and abstracts of sufficient length are retained to allow for content-based matching.

**Table 1. Number of biotechnology publications and forward citations per year (all publication matching our search key without elimination of publications having no or small abstract)**

| *YEAR* | *NUMBER OF BIOTECHNOLOGY PUBLICATIONS* | *NUMBER OF FORWARD PUBLICATION CITATIONS* | *AVERAGE NUMBER OF FORWARD CITATIONS* |
|---|---|---|---|
| 1991 | 31,381 | 1,585,560 | 50.53 |
| 1992 | 35,185 | 1,734,412 | 49.29 |
| 1993 | 38,677 | 1,913,155 | 49.46 |
| 1994 | 42,764 | 2,014,535 | 47.11 |
| 1995 | 48,092 | 2,210,601 | 45.97 |
| 1996 | 50,788 | 2,256,455 | 44.43 |
| 1997 | 53,175 | 2,441,374 | 45.91 |
| 1998 | 57,361 | 2,638,305 | 45.99 |
| 1999 | 59,866 | 2,739,699 | 45.76 |
| 2000 | 61,072 | 2,877,433 | 47.12 |
| 2001 | 62,299 | 2,697,178 | 43.29 |
| 2002 | 63,409 | 2,445,989 | 38.57 |
| 2003 | 66,564 | 2,230,473 | 33.51 |
| 2004 | 65,705 | 1,870,815 | 28.47 |
| 2005 | 72,378 | 1,609,861 | 22.24 |
| 2006 | 70,529 | 1,127,408 | 15.99 |
| 2007 | 69,756 | 738,183 | 10.58 |
| 2008 | 76,004 | 377,639 | 4.97 |
| | **1,025,005** | **35,509,075** | **34.64** |

The number of biotechnology publications in our dataset is steadily growing from 31,381 in 1991 to 76,004 publications in 2008. After a first period characterized by double-digit growth figures (from 1991 to 1995 - 10 to 12 per cent annual growth in publication outcome), we observe a period of moderate growth (4.3 to 8.0 per cent between 1996 and 1999) followed by a period of volatility during the most recent years (-2.5% to 2% with some upward outliers in 2003, 2005 and 2008).

The average number of forward publication citations (publication-to-publication citations counted by a 10-year citation window: year of publication plus following nine years)[9] for the biotechnology publications follows a more stable pattern; within a first time period observed citation rates vary between 45 and 50 (till 2000) followed by a decrease from 2000 onwards,

[9] For all forward publication citation counts in this publication we use citation counts based on a 10-year citation window except when explicitly mentioned otherwise.

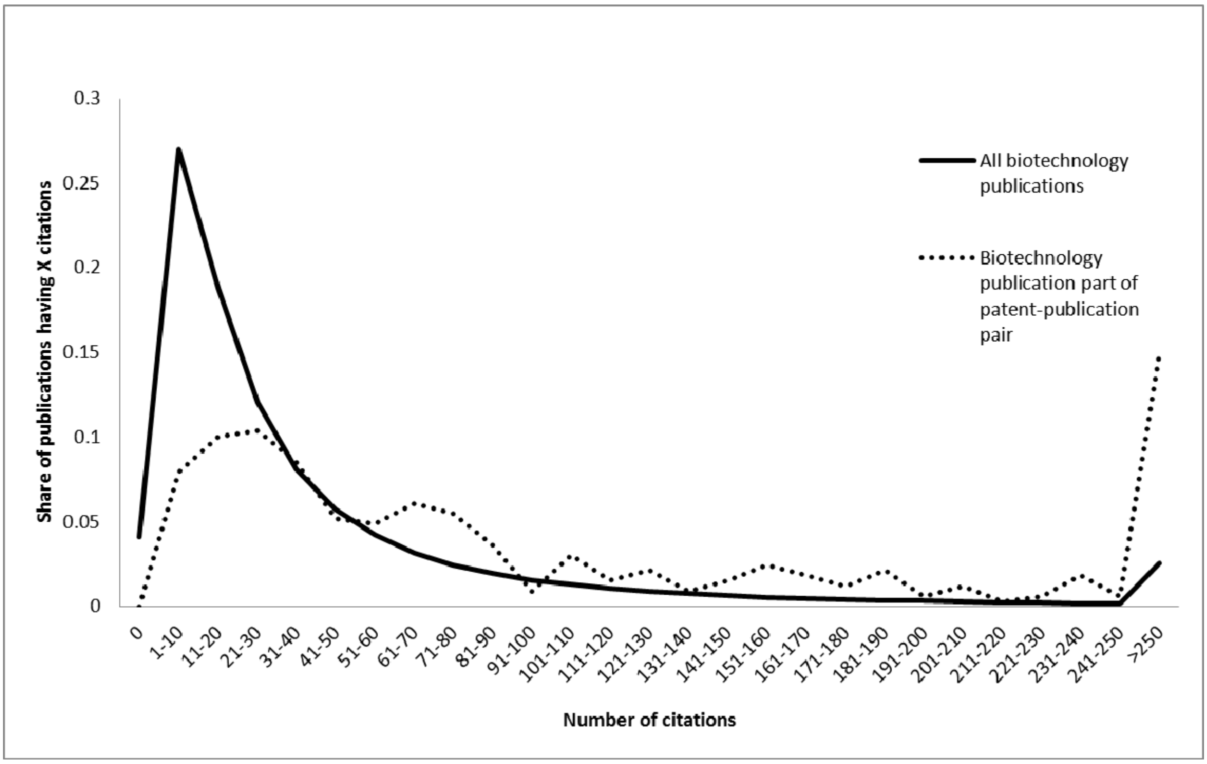
reflecting the shorter time window of observation. The average number of forward citations for all publications between 1991 and 2000 is 46.9 (median number of forward citations is 20).

For the same relevant period 1991-2000 we have 328 publications that are part of a patent-paper pair, starting from 16 in 1991 and rapidly growing to 40 in 1994, to smooth out around 40 between 1994 and 1999, and ending with a decrease to 26 in 2000. For those publications, the average number of forward citations is far more volatile throughout the years, ranging from a minimum of 77.8 forward citations on average in 1992 to a maximum of 233.9 citations on average in 1997, with no clear trend. The average number of forward citations for all publications that are part of a patent-paper pair for the total period of 1991-2000 is 161.8 (median number of forward citations is equal to 65).

On average we clearly observe substantially higher forward citation counts for publications that are part of a patent-paper pair and other publications (mean of 161.8 versus 46.9, median of 65 versus 20). But not only the average numbers are higher, the complete distribution of forward citation counts is shifted to the right in favour of publications that are part of a patent-paper pair.

Figure 1 shows the distribution of the number of forward publication citations for all biotechnology publications and biotechnology publication part of a patent-paper pair for the period 1991-2000. 25% of paired biotechnology publications have 27 or less citations compared to 7 or less citations for the first quartile for all biotechnology publications; 50% of paired biotechnology publications have 65 citations or less (20 citations for all biotechnology publications) and 75% of paired citations have 160 or less citations (48 citations for all biotechnology publications). At the right side of the distribution we observe substantial outliers, especially for publications that are related to a patent.

**Figure 1. Distribution of the number of forward publication citations for all biotechnology publications and biotechnology publications part of a patent-paper pair (1991-2000).**




One potential explanation for the higher number of forward citations for paired publications might be the difference in the number of authors. Publications having more authors tend to have more forward citations - as is confirmed by our data (an average of 38 forward citations for single authored papers up to 46 citations for publications with 5 authors and 86 citations for publications with 10 authors)[10]. We indeed observe a higher number of authors for paired publications (26% more authors on average), but this seems not to be a satisfactory explanation for the differences in citation behaviour; for publications with the same number of authors, the average number of forward citations is again substantially higher for paired publications, with a notable exception for single authored publications (an average of 19

[10] 78% of the biotechnology publications in our sample have 5 or less authors, 20% have 6 to 10 authors.

citations for single authored papers up to 135 citations for publications with 5 authors and 345 citations for publications with 10 authors)[11].

Another, more important, consideration when observing the difference in forward citation counts is the presence of a selection bias for paired publications towards "higher quality" publications. For the large overall biotechnology publication sample, all kind of quality levels will be present in the dataset. For publications that are part of a patent-paper pair, one can expect to find more publication of higher quality than average, i.e. publications valuable enough to justify costs and effort to apply for a patent. We correct this by taking into account the journal in which publications are published as an indication of the quality level of publications (i.e. we assume underlying journal impact factors are a good indication of the average quality of publications appearing in that journal).

Table 2 contains the most important journals for biotechnology publications for the period 1991-2000. The top of the table contains the most important journals in terms of the number of biotechnology publications – expressed in share of all biotech publications – while the bottom of the table contains the most important journals measured by the average number of citations for the biotechnology publications[12]. The left side of the table contains the most important journals for all biotechnology publications in our sample and the right side contains the most important journals for biotechnology publications that are part of a patent-paper pair. For every journal the average number of citations (for the biotechnology publications in our sample) and the share of biotechnology publications within our sample are listed[13].

---

[11] When comparing the number of forward citations for groups of publications with a given number of authors with a bin size of 5, paired publications always have a substantial higher number of forward citations. For publications having 1, 2, … 10 authors (the vast majority of publications), paired publications always have higher citation counts for all levels of the number of authors, except for single authored publications.

[12] The three multidisciplinary journals that were added to our selection of biotechnology patents also have a large share of all biotechnology publications in our dataset (PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA: 5.3%; NATURE, 3.1% and SCIENCE: 2.8%) but this is misleading as all publication of those journals were included in our dataset, and not only the biotechnology publications.

[13] For the right side of the table, the share of paired biotechnology publications is listed.

**Table 2. Top publishing and top cited journals for all biotechnology publications and for biotechnology publications with a paired patent (1991-2001)**

| | | ALL BIOTECHNOLOGY PUBLICATIONS | | | BIOTECHNOLOGY PUBLICATIONS WITH PAIRED PATENT | | |
|---|---|---|---|---|---|---|---|
| | | JOURNAL | AVERAGE CITATIONS | SHARE OF All BIOTECH PUBLICATIONS | JOURNAL | AVERAGE CITATIONS | SHARE OF PAIRED BIOTECH PUBLICATIONS |
| Top publishing journals | 1 | JOURNAL OF BIOLOGICAL CHEMISTRY | 67.14 | 4.96% | PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA | 108.99 | 22.26% |
| | 2 | BIOCHEMISTRY | 45.03 | 1.67% | SCIENCE | 550.68 | 8.54% |
| | 3 | JOURNAL OF BACTERIOLOGY | 34.46 | 1.66% | CELL | 366.36 | 6.71% |
| | 4 | APPLIED AND ENVIRONMENTAL MICROBIOLOGY | 33.01 | 1.59% | JOURNAL OF BIOLOGICAL CHEMISTRY | 98.67 | 6.40% |
| | 5 | BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS | 32.74 | 1.31% | NUCLEIC ACIDS RESEARCH | 57.9 | 3.05% |
| Top cited journals | 1 | NATURE REVIEWS MOLECULAR CELL BIOLOGY | 400.56 | 0.00% | NATURE | 803.13 | 2.44% |
| | 2 | ANNUAL REVIEW OF BIOCHEMISTRY | 374.20 | 0.06% | MOLECULAR CELL | 617.33 | 0.91% |
| | 3 | ANNUAL REVIEW OF CELL BIOLOGY | 305.20 | 0.01% | SCIENCE | 550.68 | 8.54% |
| | 4 | CELL | 296.03 | 0.78% | CELL | 366.36 | 6.71% |
| | 5 | ANNUAL REVIEW OF CELL AND DEVELOPMENTAL BIOLOGY | 280.95 | 0.02% | GENES & DEVELOPMENT | 307.00 | 1.52% |

## *Multivariate analysis*

To verify the significance of the observed difference when controlling for other factors multivariate analysis have been performed. Given the nature of the data (citation data) we opted for a negative binomial regression with the number of forward citations as dependent variable and a dummy variable indicating whether a publication is part of a patent-paper pair as independent variable.

To adjust for the expected difference in average quality between paired and non-paired publications (due to the potential selection bias of publications that are part of a patent-paper pair), we only include publications from journals that have at least one publication that could be paired with a patent, i.e., we only use publications that are comparable in average impact factor because they originate from the same set of journals. For this analysis, we use net citation counts, i.e. citations counts corrected for self-citations, as independent variable. This leaves 400 biotechnology publications that are part of a patent-paper pairs, and 451,803 biotechnology publications that are not part of a patent-paper pair.

We further control for journal of publications (105 distinct journals), publication document type (article, letter, note, review), number of backward publication-to-publication citations, and finally, the number of authors. We also include a time variable (1 for the first year, 1991, up to 18 for the last year, 2008) and a squared time variable to accommodate evolutions over time.

Table 3 reports the results of the regression analysis of forward publication citations of publications. Publications being part of patent-paper pairs have significantly more forward publication citations (Pair Y/N). One also notices a positive relationship between forward citations and the number of authors as well as the number of backward citations. Citation rates differ between document types: reviews receive more citations compared to articles, letters and notes. The number of forward citations differ significantly between journals (journal dummies have been included, but not reported, n=104). Finally, the observed citation rates reflect an inverse U pattern over time.

When removing outliers, i.e. all publications with a forward citation count larger than the mean plus three times the standard deviation, similar results are obtained then the ones reported in Table 3.

**Table 3. Results of negative binomial regression - Number of forward publication citations of publications (net, i.e. with exclusion of self-citations) (1991-2008)**

**Parameter Estimates**

| *Parameter* | *B* | *Std. Error* | *95% Wald Confidence Interval* | | *Hypothesis Test* | | |
|---|---|---|---|---|---|---|---|
| | | | *Lower* | *Upper* | *Wald Chi-Square* | *df* | *Sig.* |
| (Intercept) | 2.966 | .1258 | 2.719 | 3.213 | 555.643 | 1 | .000 |
| Pair (Y/N) | .450 | .0506 | .350 | .549 | 78.945 | 1 | .000 |
| Document type: | | | | | | | |
| Article | -.574 | .0113 | -.596 | -.552 | 2589.688 | 1 | .000 |
| Letter | -.774 | .0590 | -.890 | -.659 | 172.469 | 1 | .000 |
| Note | -.567 | .0175 | -.601 | -.533 | 1051.989 | 1 | .000 |
| Review | 0 | . | . | . | . | . | . |
| Number of backward publication citations | .013 | .0001 | .013 | .014 | 10416.453 | 1 | .000 |
| Number of authors | .033 | .0005 | .032 | .034 | 4613.407 | 1 | .000 |
| Time | .125 | .0015 | .122 | .128 | 7191.199 | 1 | .000 |
| Time$^2$ | -.012 | .0001 | -.013 | -.012 | 29450.994 | 1 | .000 |
| Journal dummies (n=104) | Included | | | | | | |

## *Comparison of citation counts before and after patent grant*

Inspired by the observations of Murray and Stern (2007) - a relative decline in citation patterns after patents have been granted – we verify whether the citation rates differ before and after a patent has been granted. We follow the reasoning of Murray and Stern stating that if a patent grant comes to a complete surprise to follow-on researchers, i.e. if researchers that continue working on previous discoveries are not aware of pending patent applications on those previous discoveries, a drop in citation rate can be an indication of the presence of an anti-commons effect. The reasoning behind this construct is that if researchers are not aware whether a given piece of knowledge is subject to patent filing, they will use (cite) this knowledge (publication) in a normal way. As soon as a patent covering that piece of knowledge is granted, those follow-on researchers might stop using (citing) this knowledge because of the perceived "price" (patent rights) of building on the prior discovery. Hence in case of the presence of an anti-commons effect, forward citations of publications that are part of a patent-paper pair are expected to drop as soon as the corresponding patent is granted.

To test this we split up forward citation counts for all pairs into the number of citations received before and after the grant year of their corresponding patent[14] [15] [16]. These numbers are aggregated at the level of journals and publication years, resulting in two average citations counts; one for the pre-grant period, and one for the post-grant period. Next, for every observed journal and publication year, we construct a control group that consists of all non-paired biotechnology publications published in that given journal and year. For these publications, forward citations are split up in exactly the same manner as to reflect the pre- and post-grant period. This is done as follows: if for a given journal and publication year only one paired publication is present, we split citation counts for all non-paired publications published in the same year and journal based on the lag between the publication year (journal) and grant year (corresponding patent). This is again aggregated at the level of the journal and publication year, resulting in an average citation count pre- and post-grant for non-paired patents for the given journal and publication year. If a given journal has multiple publication with a paired patent in a given publication year, we split up forward publication citation counts for the non-paired publications multiple times, once for every lag between the publication year and the grant year of the corresponding patents. All these number are aggregated at the level of the journal and publication year, resulting in an average citation count pre- and post-grant for non-paired patents. Finally, for all combinations of journals and publication years in which pairs have been observed, we calculate the ratio between citation received by pairs versus non-pairs two times: for the pre-grant period as well as the post-grant period. If an anti-commons effect would manifest itself, the ratio between pairs and non-pairs would drop significantly after granting the patent.

---

[14] For those publications linked to multiple patents (multiple members of patent families), the earliest patent grant data was used to split citations into a pre-grant and post-grant period.

[15] For this analysis, the total number of citations was used, not the net number of citations (excluding self-citations)

[16] Only publications of period 1991-2000 are included to have a full 10-year citation window for all publications and to make use of the fact that USPTO applications were not made public before 2001, making the changes of a ‘surprise’ grant to follow-up researchers more likely.

As table 4 indicates, the ratio of average citations received by pairs versus non-pairs equals to 1.71 and 1.74 before and after granting respectively. Controlling for journal and publication year, this figure means that papers that are part of a pair receive on average 71% and 74% more citations than their counterparts not belonging to a pair. While these descriptive statistics do no indicate a decline, a formal t-test reveals that both ratios are not significantly different (p=0.86). As such, our data do not show any sign of anti-common effects that become visible after patent rights have been granted.

**Table 4. Results of independent T-test – Ratio average citations pairs/non-pairs pre-grant versus post-grant (1991-2000)**

| *VARIABLE* | *CLASS* | *N* | *LOWER CL MEAN* | *MEAN* | *UPPER CL MEAN* |
|---|---|---|---|---|---|
| Ratio average citations pairs/non-pairs | Pre-grant | 288 | 1.42 | 1.71 | 2.00 |
| Ratio average citations pairs/non- pairs | Post-grant | 288 | 1.48 | 1.74 | 2.00 |
| Diff | (1-2) | | -0.43 | -0.03 | 0.36 |

| *T-TESTS* | | | | | |
|---|---|---|---|---|---|
| *VARIABLE* | *METHOD* | *VARIANCES* | *DF* | *T VALUE* | *PR > \|T\|* |
| Ratio average citations pairs/non-pairs | Pooled | Equal | 574 | -0.17 | 0.8666 |
| Ratio average citations pairs/non-pairs | Satterthwaite | Unequal | 565 | -0.17 | 0.8666 |

| *EQUALITY OF VARIANCES* | | | | | |
|---|---|---|---|---|---|
| *VARIABLE* | *METHOD* | *NUM DF* | *DEN DF* | *F VALUE* | *PR > F* |
| Ratio average citations pairs/non-pairs | Folded F | 287 | 287 | 1.29 | 0.0299 |

# Findings on citation patterns of patents (patent-to-patent citations)

Within this section we report and discuss the empirical results obtained when analysing patent citations - i.e. citations from other patent document - to patent documents that are part of a patent-paper pair. This analysis implies a comparison with patent citations to patent documents which do not belong to a patent-paper pair.

### *Descriptive results*

Table 5 provides a summary overview of the number of biotechnology patents under study as well as the observed average number of forward patent citations, organized by application year, for all biotechnology patents and for the paired biotechnology patents.

**Table 5. Number of biotechnology patents and forward citations per year (only patents with substantial abstract)**

| | ***ALL BIOTECHNOLOGY PATENTS*** | | ***PAIRED BIOTECHNOLOGY PATENTS*** | |
|---|---|---|---|---|
| ***APPLICATION YEAR*** | ***NUMBER OF PATENTS*** | ***AVERAGE NUMBER OF FORWARD PATENT CITATIONS*** | ***NUMBER OF PATENTS*** | ***AVERAGE NUMBER OF FORWARD PATENT CITATIONS*** |
| 1991 | 3,069 | 16.21 | 9 | 14.56 |
| 1992 | 3,727 | 16.14 | 11 | 24.09 |
| 1993 | 4,392 | 16.01 | 25 | 12.68 |
| 1994 | 6,170 | 14.39 | 37 | 11.16 |
| 1995 | 9,881 | 14.60 | 71 | 13.51 |
| 1996 | 5,635 | 12.13 | 33 | 6.45 |
| 1997 | 7,097 | 10.12 | 56 | 11.68 |
| 1998 | 6,974 | 8.30 | 70 | 8.84 |
| 1999 | 7,742 | 7.35 | 58 | 5.47 |
| 2000 | 7,798 | 5.46 | 65 | 3.52 |
| 2001 | 7,509 | 4.43 | 49 | 2.86 |
| 2002 | 6,315 | 3.06 | 30 | 3.17 |
| 2003 | 4,554 | 2.50 | 19 | 1.26 |
| 2004 | 3,590 | 2.16 | 23 | 11.52 |
| 2005 | 2,342 | 1.67 | 7 | 0.57 |
| 2006 | 1,170 | 1.61 | 3 | 0.33 |
| 2007 | 275 | 1.03 | 0 | N/A |
| 2008 | 8 | 0.75 | 0 | N/A |
| **TOTAL** | **88,248** | **8.94** | **566** | **8.21** |

The number of biotechnology patent grants in our dataset is starting at 3,069 patents in 1991 and exponentially growing to 9,881 patents in 1995[17]. In 1996 the number of patents falls

[17] All patent counts mentioned in this publication are granted patents by application year for patents having a substantial abstract to be of use in text mining, unless stated otherwise.

down to 5,635 to level around (and later above) 7,000 patents in the period 1997 to 2001. After 2001 the number of patents gradually diminishes.[18] The average number of forward patent citations (patent-to-patent citations) for the biotechnology patents follow a negative trend, starting around 16 from 1991 to 1993 and steadily going down from 1994 onwards (a decrease of roughly 1.5 for every year)[19] [20]. The average number of forward citations for all biotechnology patents is 8.9 (median number of forward citations is 4). For the period 1991-2000 (to allow for a sufficient time lag for citations) the average number of forward citations is 11.4 (median is equal to 5).

The number of biotechnology patents linked to a publication (566) follows a trend similar to the overall evolution of biotechnology patents: first an exponential growth phase starting from 9 in 1991 to 71 in 1995, followed by a drop to 33 in 1996 and a phase with numbers fluctuating around 63 between 1995 and 2000. Again the numbers gradually diminish after 2001, with a notable exception of 2004 (23 patents versus 19 patents for 2003 and 7 patents for 2005). The average number of forward patent citations (patent-to-patent citations) for the biotechnology patents linked to a scientific publication follow a less stable pattern and fluctuate around 13 for the period 1991-1997 (with a significant positive raise to 24 in 1992 and negative drop to 6.45 in 1996) and steadily goes down from 1997 onwards (with a steep increase to 11.52 in 2004; compared to 1.26 in 2003 and 0.57 in 2005). The average number of forward citations for biotechnology patents linked to a publication is 8.2 (median number of forward citations is 3). For the period 1991-2000, the average number of forward citations is 9.5 (median is equal to 4). These averages are about 8% lower compared to non-paired patents.

---

[18] For more recent years, trends in granted patent numbers per application year are not reliable because of declining grants due to the grant lag in patent systems.

[19] In contrast with the publication citation counts (publication-to-publication citations), patent counts in this study are not counted by a fixed citation window but continuously for all succeeding years up to 2009, the last year for which we have information available. This explains the early fall in average number of citations.

[20] The patent citation counts are corrected for patent families, both at the cited as at the citing side. At the cited side, all citations to the patent and one of its DOCDB patent family members are added together. At the citing cite, citations of multiple members of the same DOCDB patent family are counted as one.

As can be expected, patent-paper pairs are largely related to academic patenting; 52% of biotechnology patents that are linked to a publication have at least one academic patentee, compared to 18% for all non-paired biotechnology patents. Patents with at least one government or non-profit patentee are also overrepresented in the set of patents closely related to publications (23% for paired patents versus 10% for non-paired patents).

### *Multivariate analysis*

In order to assess whether observed differences are statistically significant, we performed a negative binomial regression with the number of forward patent-to-patent citations as dependent variable and a dummy variable indicating whether a patent is or is not part of a patent-paper pair as independent variable. We use all 88,248 biotechnology patents having a substantial abstract (566 patents that are part of a patent-paper pair and 87,682 patents that are not part of a patent-paper pair).

We further control for the patent system (EPO or USPTO), the number of IPC codes (technological specialization), the presence of academia as patentee, the number of backward scientific non-patent citations, the number of backward patent citations, the number of forward publication citations (citations from Web of Science publications to the particular patent), the number of inventors and the number of patentees. We also included dummy variables for all 11 biotechnology IPC subclasses (4 digits) present in our selection of biotechnology patents (see Appendix A for all IPC-codes as used in the OECD biotechnology definition). Again we include a time variable (1 for the first year, 1991, up to 18 for the last year, 2008) and a squared time variable to include the evolution over time.

Table 6 reports the results of the regression analysis of forward patent citations of patents. Patents being part of a patent-paper pairs have more forward publication citations (variable Pair Y/N), but the difference is not significant. USPTO patents have more citations than EPO patents. All other controlling variables have a significant and positive impact, except for the number of patentees, which has a negative but not significant impact and time, which displays a decreasing, curvilinear relationship with patent citations.

**Table 6. Results of negative binomial regression - Number of forward patent citations of patents (corrected for DOCDB patent family members, both at cited and citing side) (1991-2008)**

**Parameter Estimates**

| *Parameter* | *B* | *Std. Error* | *95% Wald Confidence Interval* *Lower* | *Upper* | *Hypothesis Test* *Wald Chi-Square* | *df* | *Sig.* |
|---|---|---|---|---|---|---|---|
| (Intercept) | 2.300 | .0197 | 2.262 | 2.339 | 13585.101 | 1 | .000 |
| Pair (Y/N) | .058 | .0460 | -.032 | .148 | 1.599 | 1 | .206 |
| Patent system | | | | | | | |
| EPO | -.193 | .0140 | -.221 | -.166 | 190.450 | 1 | .000 |
| USPTO | 0 | . | . | . | . | . | . |
| Subfield | | | | | | | |
| IPC subclass A01H | .114 | .0222 | .071 | .158 | 26.596 | 1 | .000 |
| IPC subclass A61K | .005 | .0094 | -.013 | .024 | .314 | 1 | .575 |
| IPC subclass C02F | -.056 | .0357 | -.126 | .014 | 2.467 | 1 | .116 |
| IPC subclass C07G | -.633 | .1145 | -.858 | -.409 | 30.574 | 1 | .000 |
| IPC subclass C07K | -.144 | .0086 | -.161 | -.127 | 282.766 | 1 | .000 |
| IPC subclass C12M | .274 | .0183 | .239 | .310 | 225.680 | 1 | .000 |
| IPC subclass C12N | .075 | .0080 | .059 | .091 | 88.167 | 1 | .000 |
| IPC subclass C12P | -.180 | .0105 | -.201 | -.160 | 292.404 | 1 | .000 |
| IPC subclass C12Q | .291 | .0095 | .272 | .309 | 941.663 | 1 | .000 |
| IPC subclass C12S | -.085 | .0473 | -.177 | .008 | 3.207 | 1 | .073 |
| IPC subclass G01N | .168 | .0096 | .149 | .187 | 305.363 | 1 | .000 |
| Number of IPC codes | .043 | .0008 | .042 | .045 | 3265.759 | 1 | .000 |
| Has university patentee (Y/N) | .036 | .0097 | .017 | .055 | 13.462 | 1 | .000 |
| Number of backward scientific non-patent citations | .003 | .0002 | .003 | .003 | 248.658 | 1 | .000 |
| Number of backward patent citations | .016 | .0003 | .016 | .017 | 3370.269 | 1 | .000 |
| Number of forward publication citations from WOS publications | .141 | .0052 | .131 | .152 | 744.627 | 1 | .000 |
| Number of inventors | .018 | .0019 | .014 | .022 | 92.591 | 1 | .000 |
| Number of patentees | -.010 | .0074 | -.025 | .004 | 1.945 | 1 | .163 |
| Time | -.063 | .0042 | -.071 | -.055 | 228.478 | 1 | .000 |
| Time$^2$ | -.007 | .0003 | -.008 | -.007 | 766.940 | 1 | .000 |

After removing outliers, i.e. all patents with a forward citation count larger than the mean plus three times the standard deviation, similar results are obtained as the ones reported in Table 6. Finally, when we limit the time period to all patents applied for between 1991 and 2000 – in order to allow all patents to have at least 10 years of forward patent citations – patent-paper pairs have less forward patent citations, but also this difference is not significant (both when including and excluding outliers). Overall, we observe no significant difference in terms of (forward) patent citations when comparing patents that are associated with a scientific publication with their solitary counterparts.

## Discussion and (intermediate) conclusions

In this paper, we have applied an advanced text mining methodology to examine the possible presence of anti-commons effects in biotechnology research. Inspired by previous work undertaken by Murray, Stern and others, we analyse citation flows stemming from patent-paper pairs present within the field of biotechnology. The delineation of the biotechnology domain was based on the use and the refined application of existing classification schemes. An elaborate text mining scheme was developed and implemented in order to identify and validate the patent-paper pairs. A total of 584 pairs were ultimately included in the citation analysis. The necessary validation and control strategies were introduced and executed. After taking into account these controls and studying the citation patterns of the documents included in the patent-paper pairs, we were not able to detect a significant anti-commons effect on the basis of the 584 pairs identified. On the contrary, scientific publications belonging to a patent-paper pair receive significantly more *scientific* citations than their counterparts for which no patent document has been identified. This difference remains outspoken (and significant) after taking into account the granted nature of implied patent documents. As such, our findings do not reveal the presence of anti-commons effects once scientific findings become translated into intellectual property rights (in this case, patents). In terms of technological citations, we observed no difference between patents belonging to a patent-paper pair and patent documents that are not associated directly with a scientific publication. As such, no additional impact – on future technological developments - is observed when patent documents are situated in the vicinity of science.

These findings add to the current stock of insights on the interaction between patenting and publication behaviour. Through the design and application of advanced text mining techniques on a broad set of data, we intended to take the current insights a step further. Extensive validation efforts were undertaken in order to confirm the results obtained.

These results definitely are an invitation to further examine the joint effects of patenting and publishing activities by scientists. The first point of attention that arises is the one of generalization towards other fields of ‘techno-scientific’ economical activity. Can we substantiate the current findings in technology domains such as materials or in other fields? The second point relates to corroborating and consolidating the robustness of the text mining methodology that was deployed, as well as a further, independent, confirmation of the optimal identification algorithm. The third point pertains to the continuous cross-validation of the results obtained with our method with the results obtained by sets of patent-paper pairs that have been constructed manually by experts.

Finally, the absence of an anti-commons effect does not imply that we have reached the end of the patent-paper debate. On the contrary, we still need a far better understanding of the many, often multidimensional, spillovers that involvement of scientists in both patent and publication activities can bring and generate. These spillovers do not only occur at the material level, but also at the immaterial, cognitive level. Understanding them and linking them to the performance of scientists in setting and advancing their research agendas, remains a question of primary importance. A better insight into these substantive relationships, both at the personal level and at the institutional level, can indeed only improve our understanding of the effective and fruitful management of scientific activity.

# References

Agrawal, A. & Henderson, H. (2002). Putting patents in context: Exploring knowledge transfer from MIT. *Management Science*, 48(1), 44-60.

Archibugi, D. (1992). Patenting as an indicator of technological innovation: a review. *Science and Public Policy*, 19, 357-368.

Argyres, N. S. & Liebeskind, J. P. (1998). Privatizing the intellectual commons: universities and the commercialization of biotechnology. *Journal of Economic Behavior and Organization*, 35(4), 427-454.

Arora, A., Fosfuri, A. & Gambardella, A. (2004). *Markets for Technology. The Economics of Innovation and Corporate Strategy.* Cambridge and London: MIT press.

Arrow K. J. (1962). *Economic welfare and the allocation of resources for invention. The rate and direction of inventive activity: economic and social factors*. Princeton NJ: Princeton University Press.

Azoulay, P., Ding, W. & Stuart, T. (2009). The impact of academic patenting on the rate, quality and direction of (public) research output. *Journal of industrial economics*, 57(4), 637-676.

Basberg, B. L. (1987). Patents and the measurement of technological change: A survey of literature. *Research Policy*, 16,(2-4), 131-141.

Baumol, W. J. (2002). *The Free-Market Innovation Machine.* Princeton NJ: Princeton University Press.

Blind, K. & Grupp, H. (1999). Interdependencies between the Science and Technology Infrastructure and Innovation Activities in German Regions: Empirical Findings and Policy Consequences. *Research Policy*, 28, 451-468.

Boitani A. & Ciciotti, E. (1990). Patents as Indicators of Innovative Performances at the Regional Level. In R. Cappellin and P. Nijkamp (Eds.), *The Spatial Context of Technological Development*. Aldershot: Avebury.

Bozeman, B. (2000). Technology transfer and public policy: A review of research and theory. *Research Policy*, 29, 627-655.

Branscomb, L. M., Kodama, F. & Florida, R. (1999). *Industrializing Knowledge: University-Industry Linkages in Japan and the United States.* London: MIT Press.

Breschi, S., Lissoni, F. & Montobbio, F. (2007). The scientific productivity of academic inventors: New evidence from Italian data. *Economics of Innovation and New Technology*, 16(2), 101-118.

Czarnitzki, D., Glänzel, W. & Hussinger, K. (2007). Patent and publication activities of German professors: An empirical assessment of their co-activity. *Research Evaluation*, 164, 311-319.

Czarnitzki, D., Glänzel, W. & Hussinger, K. (2009). Heterogeneity of patenting activity and its implications for scientific research. *Research Policy*, 38(1), 26-34.

David, P. A. (2000). The digital technology boomerang: new intellectual property rights threaten global open science. Working Papers 00016, Stanford University, Department of Economics.

Debackere, K. (2000). Academic R&D as a Business: Context, Structure and Processes. *R&D Management*, 30(4), 323–329.

Di Gregorio D. & Shane S. (2003). Why do some universities generate more start-ups than others? *Research Policy*, 32, 209-227.

Dosi, G. (2000). *Innovation, Organization and Economic Dynamics.* Cheltenham: Edward Elgar Publishers.

Etzkowitz, H. (1983). Entrepreneurial scientists and entrepreneurial universities in American Academic Science. *Minerva*, 21, 271-288.

Etzkowitz, H. (1999). Bridging the Gap: The evolution of Industry-University links in the United States. In Branscomb et al. (Eds.), *Industrialising Knowledge*. MIT Press.

Etzkowitz, H. & Leydesdorff, L. (1997). Introduction to special issue on science policy dimensions of the Triple Helix of university-industry-government relations. *Science and Public Policy*, 24(1), 2-5.

Etzkowitz, H., Webster, A., & Healey, P. (1998). Capitalizing Knowledge: New Intersections of Industry and Academia Albany. State University of New York press.

European Innovation Scoreboard (2002). Cordis Focus, December 2002.

Fabrizio, K. R. & Di Minin A. (2008). Commercialization the laboratory: Faculty patenting and the open science environment. *Research Policy*, 37(5), 914-931.

Faems, D., Van Looy, B. & Debackere, K. (2005). Interorganizational collaboration and innovation: Toward a portfolio approach. *Journal of Product Innovation Management*, 223, 238-250.

Foray, D. (2004). *Economics of Knowledge.* Cambridge, US and London, UK: MIT Press.

Freeman, C. (1987). *Technology Policy and Economic Performance.* London: Pinter.

Freeman, C. (1994). The economics of technical change. *Cambridge Journal of Economics*, 18, 463-514.

Gans, J. S. & Stern, S. (2000). Incumbency and R&D incentives: Licensing the gale of creative destruction. *Journal of Economics and Management Strategy*, 9(4), 485-511.

Hall, B., Link, A. N. & Scott, J. T. (2001). Barriers inhibiting industry from partnering with universities: Evidence from the advanced technology program. *Journal of Technology Transfer*, 26(1-2), 87-98.

Hardin, G. (1968). Tragedy of commons. *Science*, 162(3859), 1243-1248.

Heller, M. A. (1998). The Tragedy of the Anticommons. *Harvard Law Review*, 111, 621-688.

Heller, M. A. & Eisenberg, R. S. (1998). Can Patents Deter Innovation? The Anticommons in Biomedical Research. *Science*, 280,698-701.

Hellman, T. (2007). The role of patents for bridging the science to market gap. *Journal of Economic Behavior and Organization*, 63(4), 624-647.

Henderson, R., Jaffe, A. B. & Trajtenberg, M. (1996). The Bayh-Dole Act and trends in university patenting 1965-1988. *Proceedings of the Conference on University Goals, Institutional Mechanisms and the Industrial Transferabiliby of Research.*

Henderson, R., Jaffe, A. B. & Trajtenberg, M. (1998). Universities as a source of commercial technology: A detailed analysis of university patenting. *Review of Economics and Statistics*, 80(1), 119-127.

Kitch, E. W. (1977). The nature and function of the patent system. *Journal of Law and Economics*, 20(2), 265-290.

Krimsky, S. (2004). *Science and the Private Interest: Has the lure of profits corrupted biomedical research?* Rowman-Littlefield Publishing Co.

Landauer, T. K., McNamara, D. S., Dennis, S. & Kintsch, W. (Eds.) (2007). *Handbook of Latent Semantic Analysis.* Mahwah (NJ): Lawrence Erlbaum Associates.

Lester, R. K. & Piore, M. J. (2004). *Innovation-The Missing Dimension.* Harvard University Press.

Leydesdorff, L. & Etzkowitz, H. (1996). Emergence of a Triple Helix of University-Industry-Government Relations. *Science and Public Policy*, 23(5), 279-286.

Leydesdorff, L. & Etzkowitz, H. (1998). Triple Helix of Innovation: Introduction. *Science and Public Policy*, 25(6), 358-364.

Lissoni, F., Llerena, P., McKelvey, M. & Sanditov, B. (2008). Academic patenting in Europe: New evidence from the KEINS database. *Research Evaluation*, 17(2), 87-102.

Lundvall, B. A. (1992). *National Systems of Innovation: Towards a Theory of Innovation and Interactive Learning.* London: Pinter Publishers.

Magerman, T., Van Looy, B. & Song, X. (2010). Exploring the feasibility and accuracy of Latent Semantic Analysis based text mining techniques to detect similarity between patent documents and scientific publications. *Scientometrics*, 82(2), 289-306.

Magerman, T., Van Looy, B. & Baesens, B. (2011). Exploring the use of LSA/SVD text mining algorithms to map similarity between patents and publications. K.U.Leuven FBE MSI Working paper (forthcoming).

Mansfield, E. (1995). Academic research underlying industrial innovations: sources, characteristics, and financing. *The Review of Economics and Statistics*, 77(1), 55-56.

Mansfield, E. & Lee, J. Y. (1996). The modern university: contributor to industrial innovation and recipient of industrial support. *Research Policy*, 25, 1047-1058.

Merges, R. P. & Nelson, R. R. (1990). On the complex economics of patent scope. *Columbia law Review*, 90(4), 839-916.

Merton, R. K. (1968). The normative structure of science. In R. K. Merton (1979) *The sociology of science: Theoretical and Empirical Investigations.* Chicago, IL: University of Chicago Press.

Meyer, M., Sinilainen, T. & Utecht, J. T. (2003). Towards hybrid Triple Helix indicators: A study of university-related patents and a survey of academic inventors. *Scientometrics*, 58(2), 321-350.

Mowery, D. C. & Nelson, R. R. (1999). *Sources of Industrial Leadership*. Cambridge: Cambridge University Press.

Murray, F. (2006). The Oncomouse that roared: Resistance and accommodation to patenting in academic science. Sloan School of Management Working Paper.

Murray, F. & Scott, S. (2007). Do formal intellectual property rights hinder the free flow of scientific knowledge? An empirical test of the anti-commons hypothesis. *Journal of Economic Behavior and Organization*, 63, 648-687.

Nelson, R. R. (1993). *National Innovation Systems: A Comparative Analysis.* New York: Oxford University Press Inc.

Nelson, R. R. & Rosenberg, N. (1993). *Technical Innovation and National Systems*. In R. R. Nelson (Ed.), *National Innovation Systems. A comparative Analysis*: New York: Oxford University Press, Inc.

OECD (1999). University Research in Transition, OECD STI-Report OECD Publications, France.

OECD (2005). A Framework for Biotechnology Statistics, 29-32, Paris: OECD.

O'Shea R., Allen T., Chevalier, A. & Roche F. (2005). Entrepreneurial orientation, technology transfer and spinoff performance of U.S. universities. *Research Policy*, 34(7), 994-1009.

Rosenberg, N. (1998). Chemical engineering as a general purpose technology. In E. Helpman (Ed.) *General Purpose Technologies and Economic Growth.* MIT Press.

Salton, G., Wong, A. & Yang, C. S. (1975). A Vector Space Model for Automatic Indexing. *Communications of the ACM*, 18(11), 613-620.

Scotchmer, S. (2004). *Innovation and Incentives*. MIT Press.

Stephan, P., Gurmu, S., Sumell, A. & Black, G. (2007). Who's patenting in the university? Evidence from the survey of doctorate recipients. *Economics of Innovation and New Technology*, 16(2), 71-99.

Thursby, J. G. & Thursby, M. C. (2002). Who is selling the Ivory Tower? Sources for growth in university licensing. *Management Science*, 48(1), 90-104.

Thursby, M., Thursby, J. & Gupta-Mukherjee, S. (2007). Are there real effects of licensing on academic research? A life cycle view. *Journal of Economic Behavior and Organization*, 63(4), 577-598.

Trajtenberg, M. (1990). A Penny for your quotes: Patent citations and the value of innovations. *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 Technologies*, 5(1), 19-50.

Van Looy, B., Debackere K. & Andries P. (2003). Policies to stimulate regional innovation capabilities via university-industry collaboration: an analysis and assessment. *R&D Management*, 33(2), 209-229.

Van Overwalle, G. (2010). Turning Patent Swords into Shares. *Science*, 330(6011): 1630-1631.

Walsh, J. P., Arora, A. & Cohen, V. M. (2003). Science and the law: Working through the patent problem. *Science*, 299(5609), 1021-1021.

Zucker, L. G. & Darby, M. R. (2001). Capturing technological opportunities via Japan's star scientists: Evidence from Japanese patents and products. *Journal of Technology Transfer*, 26(1-2), 37-58.

# Appendix A : OECD biotechnology IPC-codes (OECD, 2005)

| IPC codes | Title |
|---|---|
| A01H 1/00 | Processes for modifying genotypes |
| A01H 4/00 | Plant reproduction by tissue culture techniques |
| A61K 38/00 | Medicinal preparations containing peptides |
| A61K 39/00 | Medicinal preparations containing antigens or antibodies |
| A61K 48/00 | Medicinal preparations containing genetic material which is inserted into cells of the living body to treat genetic diseases; Gene therapy |
| C02F 3/34 | Biological treatment of water, waste water, or sewage: characterised by the micro-organisms used |
| C07G 11/00 | Compounds of unknown constitution: antibiotics |
| C07G 13/00 | Compounds of unknown constitution: vitamins |
| C07G 15/00 | Compounds of unknown constitution: hormones |
| C07K 4/00 | Peptides having up to 20 amino acids in an undefined or only partially defined sequence; Derivatives thereof |
| C07K 14/00 | Peptides having more than 20 amino acids; Gastrins; Somatostatins; Melanotropins; Derivatives thereof |
| C07K 16/00 | Immunoglobulins, *e.g.* monoclonal or polyclonal antibodies |
| C07K 17/00 | Carrier-bound or immobilised peptides; Preparation thereof |
| C07K 19/00 | Hybrid peptides |
| C12M | Apparatus for enzymology or microbiology |
| C12N | Micro-organisms or enzymes; compositions thereof |
| C12P | Fermentation or enzyme-using processes to synthesise a desired chemical compound or composition or to separate optical isomers from a racemic mixture |
| C12Q | Measuring or testing processes involving enzymes or micro-organisms; compositions or test papers therefor; processes of preparing such compositions; condition-responsive control in microbiological or enzymological processes |
| C12S | Processes using enzymes or micro-organisms to liberate, separate or purify a pre-existing compound or composition processes using enzymes or micro-organisms to treat textiles or to clean solid surfaces of materials |
| G01N 27/327 | Investigating or analysing materials by the use of electric, electro-chemical, or magnetic means: biochemical |
| G01N 33/53* | Investigating or analysing materials by specific methods not covered by the preceding groups: immunoassay; |
| G01N 33/54* | Investigating or analysing materials by specific methods not covered by the preceding groups: double or second antibody: with steric inhibition or signal modification: with an insoluble carrier for immobilising immunochemicals: the carrier being organic: synthetic resin: as water suspendable particles: with antigen or antibody attached to the carrier via a bridging agent: Carbohydrates: with antigen or antibody entrapped within the carrier |
| G01N 33/55* | Investigating or analysing materials by specific methods not covered by the preceding groups: the carrier being inorganic: Glass or silica: Metal or metal coated: the carrier being a biological cell or cell fragment: Red blood cell: Fixed or stabilised red blood cell: using kinetic measurement: using diffusion or migration of antigen or antibody: through a gel |
| G01N 33/57* | Investigating or analysing materials by specific methods not covered by the preceding groups: for venereal disease: for enzymes or isoenzymes: for cancer: for hepatitis: involving monoclonal antibodies: involving limulus lysate |
| G01N 33/68 | Investigating or analysing materials by specific methods not covered by the preceding groups: involving proteins, peptides or amino acids |
| G01N 33/74 | Investigating or analysing materials by specific methods not covered by the preceding groups: involving hormones |
| G01N 33/76 | Investigating or analysing materials by specific methods not covered by the preceding groups: human chorionic gonadotropin |
| G01N 33/78 | Investigating or analysing materials by specific methods not covered by the preceding groups: thyroid gland hormones |
| G01N 33/88 | Investigating or analysing materials by specific methods not covered by the preceding groups: involving prostaglandins |
| G01N 33/92 | Investigating or analysing materials by specific methods not covered by the preceding groups: involving lipids, *e.g.* cholesterol |
| * Those IPC codes also include subgroups up to one digit (0 or 1 digit). For example, in addition to the code G01N 33/53, the codes G01N 33/531, GO1N 33/532, etc. are included. | |