The Light and Shade of Knowledge Recombination:

Insights from a General-Purpose Technology

CITE AS:

APPIO F.P., MARTINI A., FANTONI G. (2017), "*The Light and Shade of Knowledge Recombination: A Systematic Look at the Bioinformatics Patent Scenario*", Technology Forecasting & Social Change, Vol. 125, pp. 154-165, DOI: https://doi.org/10.1016/j.techfore.2017.07.018

1. Introduction

Bioinformatics can be thought of as a special case of important innovation as it comes at the intersection of different General-Purpose Technologies (GPTs) (Foray et al., 2009) mainly generated in the two domains of biological sciences and information and communications technology (ICT) (Majumdar et al., 2009; Lewis et al., 2016). These two domains keep on feeding each other. On one side, ICT, such as computers, networks and robotics, are nowadays playing a central role in transforming the practices of biology (Chicurel, 2002; Pop and Salzberg, 2008); indeed, these advancements are changing what biologists do, how they work, the way they carry out experiments, the universe of objects they deal with, and the kind of knowledge they can generate. They also have reoriented biology towards large-scale investigations through advanced statistical methods and big data. This change marks a break with older kinds of biological work that mainly aimed at identifying and characterizing single and specific entities.

On the other side, having seen new horizons in their research, biologists are increasingly accumulating data from multiple sources (Cook et al., 2015). Therefore, a key challenge for Bioinformatics technologies becomes the creation and maintenance of sophisticated tools and techniques for the storage, analysis, and mapping of this biological information. Therefore, it is reasonable to conceive the emergence and development of this peculiar type of GPT at the crossroads of science and technology. Such an intersection is made even

clearer by referring to the only work (Hallam, 2013) that combines historical developments and ethnographic perspectives on Bioinformatics. Herein, the author depicts a complex relationship between science and technology that contributed - and still shape – the Bioinformatics industry.

However, an important gap remains in the current state of the art: large-scale investigations on the antecedents of the patenting activity in Bioinformatics are almost silent. Pioneering contributions (e.g., Park, 2012; Rasmussen, 2010; Patel, 2003; Saviotti et al., 2000) provide a valuable but still limited view of both the emergence and evolution of the industry, not shedding light on its scientific and technological knowledge recombination and diversity. Therefore, this research focuses on exploring the interplay of scientific and technological diversity in the emergence and developments of the Bioinformatics industry. Precisely, we draw upon GPT and knowledge recombination theory to get insights on whether – and to what extent – scientific and technological diversity contribute to increase the odds for a Bioinformatics invention to have impact and become a platform for other technological developments.

This paper makes a theoretical contribution towards understanding knowledge recombination mechanisms and antecedents for GPTs, accounting for the fact that their influence on the generation of impactful inventions is nuanced. Indeed, empirical support reveals that different degrees of impact require different degrees of knowledge diversity. At the same time - and importantly for both practitioners and scholars - recombining diverse scientific and technological knowledge bases not always lead to impactful inventions. Our study may be the first revealing that for GPTs to increase their scope, inventors have to wisely recombine scientific and technological prior art. It is also one of the few studies shedding light on the relationship between diversity and recombination of the antecedents and impact of a special case of GPT.

The remaining of the paper unfolds as follows: the next section discusses relevant theory and motivates the choice of diversity indicators. Then, a discussion about research setting is provided, followed by a methodological section. Finally, results and implications conclude the article.

2. Theoretical background and research questions

Since the seminal article on GPTs (Bresnahan and Trajtenberg, 1995), this concept is gaining momentum as it has further evolved (Carlaw and Lipsey, 2011; Bresnahan, 2010; David and Wright, 2006) being applied in knowledge domains such as economic history (e.g., Crafts, 2003), industrial organization (e.g., Lindmark, 2005), economic policy (e.g., van Zon and Korenenberg, 2007), innovation studies (e.g., Andergassen et al., 2017), and translated into several models (e.g., Schultz and Joutz, 2010; Helpman, 1998). What all these approaches substantially share is the following key point: GPTs stand on technological solutions which, thanks to their high level of technological generality (Gambardella and Giarratana, 2013), may find applications in different markets and knowledge domains (Ardito et al., 2016). They are often considered as impactful innovations that pave the way to a swarm of incremental innovations transversally impacting several industries (Feldman and Yoon, 2012; Coccia, 2016). They are the engine of the economic (endogenous) growth (Ardito et al., 2016; Schaefer et al., 2014) and increased productivity in several contexts (Feldman and Yoon, 2012). Overall, GPTs give birth to technological discontinuities which, for their inherent general scope, bring with them a broad array of applications.

A definition which gathers together these perspectives is provided by Lipsey and co-authors (2005, p. 96): a GPT "*is a single technology, recognizable as such over its whole lifetime that initially has much scope for improvement and eventually comes to be widely used, to have many uses, and to have many spillover effects*." Still Lipsey and co-authors (2005) distinguish between five classes of GPTs namely, materials, ICT, power sources, transportation equipment, and organizational forms.

But Bioinformatics escapes univocal identifications: it does not seem to fit exactly in one of the five classes only and this is why it is identified as a special case of GPT (Foray et al., 2009). To give the reader an idea of the wide spectrum Bioinformatics opens, it suffices to consider some examples of the ways it aids the experimental approaches: first, for bridging proteins, DNA, and RNA sequences, a number of databases have been built (e.g., NCBI's SNP database[1], ERGO 2.0 genome browser[2]) for which BLAST[3] (Basic Local Alignment Search Tool) can be used; second, for searching for functional patterns in proteins and nucleic acids, the ExPASy[4] (Expert Protein Analysis Model), or the ENSEMBL[5] tool; finally, GPCRDB[6] is usually used for determining if there are known interactions among proteins. Bioinformatics technologies, applications, and solutions touch the following fields: microbial genome applications, molecular medicine, personalized medicine, preventative medicine gene therapy, drug development, biotechnology, climate change studies, insect resistance, vetinary science, evolutionary studies.

Notwithstanding its revolutionary role, and although the existence of a rich theoretical literature on GPTs (Feldman and Yoon, 2012), sound empirical investigations on the antecedents of GPTs are scant (Thoma, 2009; Lo and Sutthiphisal, 2010). As recent literature advocates, inventions of such a kind may require a broad search for information and the recombination of different knowledge bases (Kaplan and Vakili, 2014). Recombination mechanisms are increasingly being associated with the accomplishment of impactful inventions (Fleming and Sorenson, 2004; Schilling and Green, 2011; Carnabuci and Operti, 2013): in fact, recent research shows it is going through a variety of cognitive-search processes (Acar and Van den Ende, 2016) that atypical (Uzzi et al., 2013) or unconventional combinations (Simonton, 1999) make concrete. The fact that there is a link between

[1] https://www.ncbi.nlm.nih.gov/SNP/
[2] https://www.igenbio.com/
[3] https://blast.ncbi.nlm.nih.gov/Blast.cgi
[4] http://www.expasy.org/
[5] http://www.ensembl.org/index.html
[6] http://gpcrdb.org/

recombination of diverse knowledge bases and impact is increasingly recognized and empirical studies start to find systematic evidence on it (Keijl et al., 2016). Developing technologies drawing on diverse knowledge areas contribute to the creation of GPTs, since this enhances inventions' technological generality. Accordingly, we take the view of inventions as the outcome of a recombination process over technology and scientific landscapes (Fleming and Sorenson, 2001; Savino et al., 2015, Nakamura et al., 2015). This recombination of diverse knowledge bases, if not their interplay, may play an important role in spanning the boundaries of GPTs (Novelli, 2015). According to Petsas (2003), the inputs of GPTs are basically complementary and it makes relevant looking at whether – and to what extent – diversity is at the basis of the impact of the Bioinformatics industry.

To this purpose, we will have a deep look at the patents' prior art (Trajtenberg et al., 1997; Chen and Hicks, 2004; Callaert et al., 2006) as they reflect the recombined knowledge at the basis of inventions in the field of investigation. Prior art in patents can be distinguished in patent references - PRs, representing the cited technological contributions - and non-patent references - NPRs, representing the cited scientific contributions. Analyzing their inherent diversity may provide us with the opportunity to systematically examine the relationships between scientific and technological contributions at the basis of key inventions in the Bioinformatics industry.

Former contributions (Park, 2012; Rasmussen, 2010; Patel, 2003) focus on patents in Bioinformatics and provide valuable insights when it comes to consider the identification of the main technological classes and subclasses. However, they have two main limitations: first, they do not focus on the characteristics of prior art; second, they do not provide systematic evidence of the link between diversity in the prior art and GPT impact. Following our theorizing, two main research questions emerge:

What is the link between knowledge recombination and the impact of Bioinformatics inventions? And, to what extent does diversity of scientific and technological knowledge bases matter?

3. Characteristics of Bioinformatics

3.1 Early days and growth of the industry

The history of Bioinformatics traces back to the late 1970s when Staden, in a series of studies published on *Nucleic Acids Research* (seminal paper published in 1977), outlined the basics for the development of computer programs allowing researchers to analyze DNA sequences. However, they could be used on large mainframe computers and run with non-standard programming languages; these two characteristics made adapting them to small (48K) microcomputers tile consuming and not even economically affordable. These reasons, along with the need to facilitate dissemination, inspired the work of James M. Pustell (then Ostell) at the Kafatos Laboratory in Harvard's Department of Cellular and Development Biology. He worked on – and improved – four aspects: interaction, data management, data analysis, and software compatibility (Pustell and Kafatos, 1982). It was a success (Hallam, 2013) and since its inception, the Bioinformatics industry underwent periods of technological and scientific challenges, feeding each other over a long period of time (Ouzonis and Valencia, 2003).

What follows is an overview of the challenges addressed by technological advancements. In the 1990s the most urgent challenge was to deploy systems able to collect huge amounts of data (Mushegian, 2011). At that time, the technical advancements were as such that tools like the Genome Sequence Database allowed scientists to process data within 13 hours; a performance that only few years before was of 13 months (Robbins, 1996). Since the beginning of the Human Genome Project in 1990, always more sequence data entered the databases (Robbins, 1996).

Then, a new challenge showed up: data integration. As Robbins (1996, p. 468) puts it: "*sequence data had to be linked to map data; protein structures, connected to metabolic function; species data, connected with ecosystem data*." Although progress has been made, still today this challenge is considered as perennial issue in that the increase in the quantity of data brings complexity (Goble and Stevens, 2008). Different approaches to data integration are nowadays available ranging from service-oriented architectures, data warehousing, view integration; workflows, mashups, and semantic web (Goble and Stevens, 2008).

Another set of challenges laid in the problem of data standardization and data indexing (Robbins, 1996). Recent technological advancements allow for solutions which can also allow standardizing and indexing big data, mainly based on supervised and unsupervised machine learning techniques (Greene et al., 2014).

In terms of scientific developments, it is worth considering that it is only from the mid-1990s onward that new journals, new conferences, and new training programs (and textbooks) began to appear supporting this new domain of knowledge (Hallam, 2013): in 1994, the first statement on the birth of computational biology as a discipline in its own right was made by the new *Journal of Computational Biology* (Davison, 1994); in 1995, the first Bioinformatics textbook was published Waterman (1995); in 1998, the journal *Computer Applications in the Biosciences* (established in 1985) changed its name to *Bioinformatics* (Hallam, 2013) highlighting "the importance of Bioinformatics, a subject which is emerging as one of the leading scientific disciplines for the next century" (CABIOS, 1997).

Afterwards, the need for Bioinformatics specialized curricula emerged (Altman, 1998). Under the headline '*Bioinformatics: Jobs Galore*', Science Careers reports in 2000 that "everyone is struggling to find people with the Bioinformatics skills they need" (Taylor, 2000). Public entities in general and particularly universities, started to pay attention to the rise of this new discipline (Black and Stephan, 2004). In the period 1997-2004, the number

of Bioinformatics programs within universities more than tripled, twelve professional science Master's in Bioinformatics were created, the number of Doctoral programs more than doubled, whereas the number of Bachelor's programs tripled.

Scientific projects like the Human Genome Project (HGP) strongly contributed to institutionalize Bioinformatics (Hallam, 2013). By the time the HGP came on the scene, independent initiatives spurred the growth of new institutional spaces for data work. What the Congress reasoned about the creation of the National Center for Biotechnology Information (NCBI) (Cook-Deegan, 1994) was that "*knowledge in the field of biotechnology is accumulating faster than it can be reasonably assimilated*" and that advances in computation were the solution (Smith, 2008). Specifically, the Congress emphasized the need of a distinct institutional home to deal with the different skills needed to design, development, implement, and manage biological information (Hallam, 2013). For the same reasons, in Europe, the European Bioinformatics Institute (EBI) was established as a quasi-independent outstation of the European Molecular Biology Laboratory (EMBL). The acceleration and rapid completion of the HGP made it clear that, as one Nature editor put it, "like it or not, big biology is here to stay" (Editorial Nature, 2001, p. 545). Bioinformatics and HGP fed each other by enabling the production and synthesis of more and more data.

2.2 Bioinformatics market and patents

Already in 2000, the combination of segments characterizing Bioinformatics exceeded an estimated $700 million, making Bioinformatics one of the fastest growing areas of all life science related markets (SDI, 2001). Recent analyses confirm this picture: the global Bioinformatics market has been valued at $3.5 billion in 2013 and it is poised to reach $12.5 billion by 2020 at a Compound Annual Growth Rate (CAGR) of 19.49% (Research and Markets, 2015). These numbers are probably underestimating the fact that certain pharmaceutical and biotechnology firms have extensive in-house Bioinformatics expertise, which never offers its services for sale (Saviotti et al., 2000). In any case - and consequently

- the most active companies understood the impressive economic potential of this industry and started aggressively seeking patent protection for their Bioinformatics inventions with some attempting to corner the market (Vorndran and Florence, 2002; D'Souza, 2004; Rasmussen, 2010).

However, patenting in Bioinformatics is complex as it deals with many dimensions such as lines of code, algorithms, data content, structures and user interfaces (Rasmussen, 2010). Despite some unique legal problems of this discipline (Ostler and Gollin, 2015; Welch, 2002) and the persisting differences among the legal systems (McCubbin, 2003), patents are of greatest assistance in protecting most forms of intellectual property associated with Bioinformatics (Harrison, 2003). Figure 1 shows the evolution of the patenting activity in the Bioinformatics industry and relates it to the scientific and institutional milestones introduced in the former section.

FIGURE 1 HERE

Figure 1. Number of patents over the years and milestones

As we will show in the next section, what further complicates the picture is that by entailing Bioinformatics a combination of GPTs, inventors potentially may come from different domains, relying then on diverse knowledge bases (Hallam, 2013).

3. Methodology and data set

3.1 Collecting Bioinformatics patents

As there is no single and unique classification for Bioinformatics patents, a series of steps have been undertaken to carry out preliminary patent-based analyses by relying on the United States Patent and Trademark Office (USPTO)[7] information. This allowed us to expand results stemming from former studies (Park, 2012; Rasmussen, 2010; Patel, 2003). The following data search and collection routine has been implemented:

[7] http://www.uspto.gov/

- First large-scale extraction of patents belonging to USPTO sub-classes identified in Park (2012) which in turn are based on Rasmussen (2010) and Patel (2003). First dataset generated.
- Multiple keywords search based on Hallam (2013) (e.g., 'Bioinformatics', 'computational biology', 'genomics', 'data-based biological knowledge') relying on patent title, abstract, and claims of patents to include only those which satisfied both the biotechnology and information technology content of Bioinformatics. Second dataset generated.
- Merging of the two samples coming from the two former points.
- Semi-automatic review and refinement of selected patents (including patent description), to ensure consistency.

This routine resulted in a database of 14,754 Bioinformatics patents in the 1976-2014-time window. It is worthy to notice that 12% of the patents come from individual assignees (Appendix 1). From the distribution of the NAICS codes, we can also observe that the remaining assignees come from diverse industries and this confirms the fact that Bioinformatics is increasingly becoming an attractive field for heterogeneous actors. Universities are playing an important role too.

3.2 Extracting prior art information

Of all the Bioinformatics patents, we collected both cited non-patent references (NPRs) and patent references (PRs). Some problematic issues may exist in referring to patent citations as a way to identify knowledge linkage and recombination (Chen, 2007): the first disputation entails the difference between patent citations added by the applicant and examiners (e.g., Alcacer et al., 2009); second, some scholars point out that only NPRs self-cited by the inventors should be used as a way to trace the link between science and technology, as the ones from examiners and non-self-citations by the inventor introduce noise (e.g., Li et al., 2014). Despite these controversial issues, by using patent citations (both NPRs and PRs),

scholars have revealed knowledge diffusion and transfer (e.g., Nelson, 2009), trajectories and paradigms (Érdi et al., 2013); all in all, they provided with a more systematic conceptualization of innovation dynamics and related policies by building more accurate citations-based indicators (e.g., Callaert et al., 2006; 2012). We then follow Chen (2017) who argues it is reasonable to assume that patent citations are linked to knowledge.

Several stages of data cleaning and harmonizing followed in order to extract the essential information allowing us to calculate indicators. A database has been built in which the following information showed up: patent ID; filing date; NPRs string; PRs string. This routine resulted in a database of 397,746 NPRs in the 1974-2013-time window and 194,288 PRs in the 1972-2013-time window.

Concerning the NPRs, the sources have been extracted from the entire string by creating specific macros in VBA Project. Appendix 2 provides the reader with an overview of the main NPRs. No one of the Bioinformatics journals show up in the top 1%, the reason is twofold: first, the Bioinformatics field is still in development and relatively recent journals need time to become relevant; second, it is an obvious consequence of the interdisciplinary nature of the field. Appendix 3 reports the main Bioinformatics journals and their frequencies, accounting for a 1.5% of the NPRs.

Also, the sources needed to be classified as Science (SCI-covered) or Not Science (Not SCI-covered) by following Callaert and co-authors (2006). To do that, we matched data in the Journal Citation Report's database[8]. The matching was made on the basis of Journal title, and impact factor. We uniquely allocated 365,885 (91.98%) NPRs[9] (see Table 1).

TABLE 1 HERE

Table 1. NPRs classification

[8] http://thomsonreuters.com/en/products-services/scholarly-scientific-research/research-management-and-evaluation/journal-citation-reports.html

[9] Complete list available upon request.

These values indicate that the sources that scientific journals are the main source to the development of Bioinformatics innovations are mostly. Figure 2 shows their distribution over the years.

FIGURE 2 HERE

Figure 2. Number of NPRs (green), Non-Journal (blue), SCI-covered (purple) and Not SCI-covered (red) references over the years (log scale)

A relevant percentage of non-scientific items is present thus signaling that books, manuals, patent related documents, database entries, magazines (to mention just the main categories) have contributed to provide the necessary knowledge to bring the innovations into this sector. This is a first symptom of the necessary diversity of sources at the basis of the Bioinformatics inventions.

Concerning the PRs, we extracted the technological sub-classes (2 and 3 digits) by following the IPC classification[10]. Table 2 shows the main technological sub-classes cited into Bionformatics patents[11]. This preliminary information confirms the interdisciplinary nature that lays at the basis of the growth and development of the Bioinformatics industry.

TABLE 2 HERE

Table 2. PRs classification

3.3 Indicators

A diversity index was selected to account for quantity and quality of both science (NPRs) and technology (PRs) in patents. Starting from a general measure of information content called Renyi entropy R_i of order α, we got the value of interest for this study which is the limit of the equation (Hill, 1973; Pielou, 1975) when α approaches 1. This yields to:

$$H' = R_i = - \Sigma_i \, p_i \ln p_i$$

[10] http://www.wipo.int/classifications/ipc/en/

[11] Complete list available upon request.

where p_i is the proportion of individuals found in the *i*-th species. This is the well-known Shannon-Wiener measure of species diversity (H'), which is widely used in a number of different areas (Stirling, 1998; McDonald and Dimmick, 2003; Margaleff, 1972). The Shannon-Wiener diversity index H' is used to calculate the NPRs and PRs diversity.

In order to assess the odds for a Bioinformatics invention to have impact, the number of Forward Citations (FCs) is used as a proxy. Indeed, the number of FCs can be thought of as signalling the importance of a certain technology (Albert et al., 1991; Carpenter et al., 1993; Jaffe et al., 2000), especially in the case of GPTs (Youtie et al., 2007). The extent to which a GPT is cited and adopted for the development of future inventions gives an indication of their importance and usefulness (Keijl et al., 2016). A number of studies also support the evidence that FCs are a good indicator of both the social (Trajtenberg, 1990; Sapsalis et al., 2006) and private value (Harhoff et al., 1999; Lemley and Shapiro, 2005; Gambardella et al., 2008) of specific inventions. Another interesting characteristic stems from the fact that the distribution of FCs is highly skewed with a large share of patents receiving no citations and a small minority of patents with a large number of FCs (Arts et al., 2013). Linking this skewedness to the value of an invention allows us to argue that outliers in the distribution of FCs pertain to more important inventions (Arts et al., 2013), with breakthrough patents falling in the top 1% or 5% in terms of FCs (Ahuja and Lampert, 2001; Singh and Fleming, 2010).

3.4 Model

We run a series of logistic regression models (Hosmer et al., 2013) in order to assess the likelihood of being highly cited (through the number of FCs at 1, 2, and 5 Standard Deviations).

$$\Pr(Y=1|X_1, \ldots, X_8) = \beta_0 + \beta_1 * H'_{NPRs} + \beta_2 * H'_{PRs\_class} + \beta_3 * H'_{PRs\_subclass} + \beta_k * Controls$$

with k=4, …,8. Table 3 shows a summary of the variables used into the model.

TABLE 3 HERE

Table 3. Variables for logistic regression models

Figure 3 summarizes the method and data collection process:

FIGURE 3 HERE

Figure 3. Flowchart to summarize method and data collection process

4. Results

4.1 Descriptive Statistics

In terms of correlations (see Appendix 4), we may notice that diversity at the level of NPRs has a negative correlation with all the levels of impact we are going to use. Even though the coefficients are not that large, they are still highly significant. NPRs diversity's effect is lower for impactful innovations at the average level of FCs (-.018, $p<.05$), reinforces for those at 1SD (-.076, $p<.001$) and 2SD (-.073, $p<.001$), lowering again for extreme cases at 5SD (-.054, $p<.001$). Instead, a positive and significant correlation emerges for PRs regardless the fact their diversity is at the level of technological classes or subclasses. However, higher coefficients emerge when it comes to consider the correlation between the diversity at the level of subclasses and the different levels of impact (e.g., .206, $p<.001$ for 0SD cases; or .116, $p<.001$ for 2SD cases). This is supposedly due to the hierarchical structure of the IPC classification (Eisinger et al., 2013). As a consequence, the measure of diversity at one level does not take into account diversification at the underlying levels. This difference may also be reflected in the opposite and significant correlations between NPRs and PRs diversity (both at the level of classes and subclasses). Overall, all the variables pertaining to the scientific knowledge bases seem to run against the odds to generate impactful innovations in the Bioinformatics industry; the opposite holds for the technological knowledge bases.

4.2 Gaining impact by recombining science and technology?

In Table 4 we compare different logistic regression models, according to the definition of impact. The regression model with the interaction effects shows that the interplay of NPRs diversity and PRs diversity comes along with a small (but significant) decrease in the odds of generating impactful inventions. Such decreases become not significant as soon as extreme events are taken into account: for events falling in the 5SD of the FCs distribution, it seems that the interplay of scientific and technological knowledge bases does not play a significant role. On one side, recombining diverse knowledge bases at the level of NPRs and PRs technological subclasses has a negative impact only on the cases at 0SD and 1SD. On the other side, the negative effect of recombining NPRs knowledge bases and PRs technological classes expands until cases at 2SD.

NPRs diversity shows three patterns. First, it has a positive coefficient when cases at the average level of SD are considered. Second, for cases at 2SD it decreases the odds of getting impactful inventions and in a significant way as soon as there is no interplay with PRs diversities. Third, and regardless the presence of interaction effects, it is still negative but not significant for more extreme cases namely, 2SD and 5SD.

TABLE 4 HERE

Table 4. Logistic regression model results (coefficients reported)

PRs diversity shows a constant pattern: it has positive, always significant, and reinforcing coefficients at both the level of classes and subclasses. However, PRs diversity at the level of technological subclasses exhibits higher coefficients only for impact at the average level. Always higher coefficients for PRs diversity at the level of technological classes can be observed for increasingly extreme cases.

By assessing the interaction effects between the main predictors, different optimal strategies may be identified (see Table 5).

TABLE 5 HERE

Table 5. Interaction effects assessment

The overall effect of NPRs diversity on the odds of generating impactful inventions in the Bioinformatics industry is positive when it comes to consider events at the average level of FCs (0SD); it ranges from -4.4% to +12.6%. The maximum level of increase (+12.6%) occurs when there is no interaction effect with the two PRs diversity indicators. The lower level instead occurs either when NPRs diverse knowledge bases are recombined with PRs diverse technological classes (-4.4%). Another interesting emerging aspect - which also reflects the results of the regression models - is that moving away from cases at the average level of FCs (0SD) implies that the overall effect of NPRs diversity is either negative or not significant. Furthermore, and although the magnitude is not so high (from -4.4% to -1.4%), recombination while shifting towards more extreme cases does not help in gaining impact. For extreme cases (1SD, 2SD, and 5SD), the overall effect of NPRs diversity is not significant.

PRs diversity at the level of technological subclasses has a positive effect on the odds of generating impactful inventions with a magnitude ranging from +4.4% to +10.7%. For cases at 0SD and 1SD the PRs diversity at the level of subclasses allows for higher increase in the odds of generating impactful inventions (+8.7% and +5.9%). The interaction effect with NPRs diversity is still a good option (+8.7% and +4.4%). Instead, for more extreme cases (at 2SD and 5SD), either by recombining with diverse NPRs or not, there is a positive and significant increase in the odds (from +8.9% to 10.7%). An increase in the odds is allowed as soon as inventors decide to leverage on the diversity of PRs' technological subclasses.

PRs diversity at the level of technological classes has overall positive influence on the odds with a magnitude ranging from +5.9% to 18.3%. This predictor shows better performance when there is no interaction effect with NPRs diversity, and that happens when there is no recombination of technological and scientific knowledge bases. For increasingly extreme events (at 2SD and 5SD), the values are higher ranging from +8.5% to 18.3%.

5. Discussion and Implications

In this paper, we examined whether – and to what extent – scientific diversity and technological diversity contributed to the rise and development of the Bioinformatics industry. We analyzed the Bioinformatics patented inventions (USPTO) in the period 1976-2014 and carried out a systematic assessment of their prior art (NPRs and PRs). By identifying scientific knowledge bases with NPRs and technological knowledge bases with PRs, we calculated their diversity by means of an entropy-based indicator namely, the Shannon-Wiener diversity index. We then performed logistic regressions in order to assess whether the diversity of NPRs and PRs might be considered good predictors of inventions' impact. This analysis was done by considering different degrees of impact taking into account the characteristics of the FCs distribution.

From the amount of both NPRs and PRs, and by looking at historical evidence of the development of the Bioinformatics industry, we preliminary argued that both science and technology play a role. In terms of quantity, the number of cited NPRs more than doubles the number of PRs. However, their contribution is somewhat different. Indeed, when we dug into the types of NPRs, we noticed the contribution of both SCI-covered references (e.g., indexed academic journals) and their non-scientific counterparts (e.g., books, patent related documents, and database accessions) were relevant. Also, the fact that most of SCI-covered sources are not Bioinformatics journals provided preliminary signs that diversity may matter. A similar trait emerges when we look at the technological classes and subclasses of PRs. Overall, in terms of quantity, Bioinformatics patents lay on a higher number of NPRs than PRs.

When it comes to consider their diversity, descriptive statistics showed a negative correlation between NPRs diversity and all the levels of impact considered. On the contrary, and although with slightly different coefficients, PRs diversity is positively correlated with all the impact measures we used. Overall, diversity (or quality) of cited NPRs does not seem to

be as important as quantity to the development of important inventions in the Bioinformatics industry.

In our logistic regression models, a more nuanced picture emerged. First, by considering the models without interaction effects, NPRs diversity has a positive and significant effect on the odds of generating inventions at the average level of FCs; then, its effect becomes negative and highly significant concerning the inventions at 1SD of FCs giving inventors a decrease in the odds which amounts to -19.9%; finally, for extreme cases (2SD and 5SD in the FCs distribution), the coefficient is no longer significant. The coefficient concerning the PRs diversity at the level of technological classes and subclasses exhibited a positive sign and significantly contribute to increase the odds of generating impactful inventions. PRs diversity matters especially when dealing with extreme cases (at 5SD). But to escape the average impacts, inventors should leverage on PRs diversity at the level of technological classes in that an 18.3% increase in the odds can be reached out.

Second, by considering the logistic regression models with the interaction effects between NPRs diversity and PRs diversity, results indicated another interesting aspect. The interplay of scientific and technological knowledge has almost always a negative coefficient, although its significance shades as soon as extreme cases are considered. Recombining science and technology does not seem to be the best option to increase the odds to gain impactful inventions in the Bioinformatics industry.

This research moves a step further in the understanding the determinants of GPTs. Analyses of the patenting activity of the Bioinformatics industry, although valuable, only provide with evidence about the main technological classes of patents. To the best of our knowledge, we for the first time open the black-box of the Bioinformatics patents in order to shed light on the quantity and diversity (or quality) of the cited NPRs and PRs. Understanding how the latter are recombined is of fundamental importance in that insights are needed to disentangle the emergence and development of GPTs.

Our findings contribute and extend the existing empirical and theoretical debate on knowledge recombination (Carnabuci and Operti, 2013; Gruber et al., 2013; Blackwell et al., 2009; Fleming and Sorenson, 2004), especially when it claims the higher the recombination of knowledge from areas distant to the one where someone is innovating, the higher the chances of generating important innovations. However, we highlight that only recombination occurring within scientific domains or technological domains can be considered, as their interplay may be detrimental to gain impactful inventions. Also, our contribution shows that the different degrees of impact require different degrees and types of knowledge recombination either within scientific domains or technological ones. NPRs diversity is more relevant for those inventions having a low impact; on the contrary, PRs diversity has impact for cases having moderate to extreme impact.

Our findings also have important practical and policy implications. First, inventors in GPTs have to wisely recombine existing and diverse scientific/technological knowledge bases in order to increase the chances to generate impactful inventions. Within the boundaries of the considered industry (Bioinformatics) we can claim that it is better for inventors to rely either on diverse scientific knowledge bases or diverse technological knowledge bases. Their interaction does not represent the best recipe. Second, inventors (and the organizations they belong to) may reinterpret our findings in the light of the framework Murmann and Frenken (2006) advanced, perhaps locating Bioinformatics inventions in the radical-square innovation quadrant: a large scope of new knowledge mainly residing in PRs diverse sources increases the odds of getting high performance improvements. Third, as recombination of diverse PRs seems to be the strongest mechanism to gain impact, by leveraging on it inventors in the Bioinformatics industry could make the transition from technological trajectories to new paradigms potentially more concreate (Dosi, 1982). Policy makers may find interesting the fact that understanding what is behind the scene of GPTs, they can foresee better ways to prioritize investments for assuring a long-term economic growth.

Also, they may find ways to further stimulate collaborative efforts between the heterogeneous actors to enact network dynamics, enabling them to better recombine their knowledge bases. Finally, by opening the black box of the NPRs and PRs, policy makers can start posing the basis for a more univocal way to identify Bioinformatics patents in the different patent classification systems.

6. Limitations and Future Research

This research does not come without limitations. First and foremost, we relied on USPTO patents only. Future research may extend their analyses by including European patents as well in order to assess whether some difference exist in both the PRs and NPRs citations patterns. Second, we do not dig into the content of the cited prior art; in this vein, future studies may investigate the semantic content of prior also by looking at their evolution. Therefore, new content-based indicators signaling impact of Bioinformatics inventions can be generated and tested. Third, our focus is on a peculiar case of GPT which may not allow for generalizing our results. Future researchers may further elaborate on this topic by comparing other GPTs and questioning or supporting our findings. Fourth, we do not dig into the inventor's characteristics; this aspect may be useful to understand who the most prominent scientists in the Bioinformatics field are, who among them is able to play the ambidextrous role of producer of scientific and technological outcomes, or how their curricula look like. This may provide future readers with a complete picture of the necessary skills to survive complex industries like Bioinformatics. Fifth, studying the impact generated by networks of inventors and whether these networks are characterized by strong industry-university links is worthy of further investigation. Finally, analyzing how – and to what extent – the impact of bioinformatics patents varies across industrial boundaries (Messeni Petruzzelli et al., 2015) represents an interesting research avenue.

Acknowledgements

We thank Dr. Riccardo Apreda (ErreQuadro) and Eng. Giordano Campatelli for their support in the data collection process.

References

O.L. Acar, J. Van den Ende, Knowledge distance, cognitive-search processes, and creativity: the making of winning solutions in science contests, Psychol. Sci. 27(5) (2016) 692 699.

G. Ahuja, C.M. Lampert, Entrepreneurship in the Large Corporation: A Longitudinal Study of How Established Firms Create Breakthrough Inventions, Strategic Manage. J. 22(6/7) (2001) 521 543.

M.B. Albert, D. Avery, F. Narin, P. McAllister, Direct Validation of Citation Counts as Indicators of Industrially Important Patents, Res. Policy 20(3) 1991 251 259.

J. Alcacer, M. Gittelman, B. Simpat, Applicant and examiner citations in US patents: an overview and analysis, Res. Policy 38(2) (2009) 415 427.

R.B. Altman, Editorial: A Curriculum for Bioinformatics: The Time is Ripe, Bioinformatics 14(7) (1998) 549 550.

R. Andergassen, F. Nardini, M. Ricottilli, Innovation diffusion, general purpose technologies and economic growth, Struct. Change Econ. Dynam. 40(1) (2017) 72 80.

L. Ardito, A. Messeni Petruzzelli, V. Albino, Investigating the antecedents of general purpose technologies: a patent perspective in the green energy field, J. Eng Technol. Manage. 39(January/March) (2014) 81 100.

F.P. Appio, A. Martini, A. Messeni Petruzzelli, P. Neirotti, B. Van Looy, Search mechanisms and innovation: an analysis across multiple perspectives, Technol. Forecast. Soc. Early view (2017).

S. Arts, F.P. Appio, B. Van Looy, Inventions shaping technological trajectories: do existing patent indicators provide a comprehensive picture? Scientometrics 97(2) (2013) 397 419.

G.C. Black, P.E. Stephan, Bionformatics: recent trends in programs, placements, and job opportunities, Report to the Alfred P. Sloan Foundation (2004). Available at: https://www.calstate.edu/csuperb/faculty/resources/documents/reportbioinfsloan_june04.pdf (Accessed on November 2015).

A.F. Blackwell, L. Wilson, A. Street, C. Boulton, J. Knell, Radical innovation: crossing knowledge boundaries with interdisciplinary teams. Technical Report, University of Cambridge, Computer Laboratory, n.760 (2009). Available at: https://www.cl.cam.ac.uk/techreports/UCAM-CL-TR-760.pdf (Accessed on March 2016).

T. Bresnahan, General purpose technologies, in: Handbook of the Economics of Innovation, B.H. Hall, N. Rosenberg (Eds.), Elsevier, (2010) 761 791.

CABIOS, Bioinformatics online (1997). Available at: http://bioinformatics.oxfordjournals.org/content/13/6/local/front-matter.pdf (Accessed on February 2016).

J. Callaert, B. Van Looy, A. Verbeek, K. Debackere, B. Thijs, Traces of Prior Art: An analysis of non-patent references found in patent documents, Scientometrics 69(1) (2006) 3 20.

J. Callaert, J. Grouwels, B. Van Looy, Delineating the scientific footprint in technology: identifying scientific publications within non-patent references, Scientometrics 91(2) (2012) 383 398.

K.I. Carlaw, R.G. Lipsey, Sustained endogeneous growth driven by structured and evolving general purpose technologies, J. Evol. Econ. 21(4) (2011) 563 593.

G. Carnabuci, E. Operti, Where do firms' recombinant capabilities come from? Intraorganizational networks, knowledge, and firms' ability to innovate through technological recombination, Strategic Manage. J. 34(13) (2013) 1591 1613.

M. Carpenter, F. Narin, Validation study: patent citations as indicators of science and foreign dependence, World Pat. Inf. 5(3) (1993) 180 185.

L. Chen, Do patent citations indicate knowledge linkage? The evidence from text similarities between patents and their citations, J. Informetr. 11(1) (2017) 63 79.

C. Chen, D. Hicks, Tracing knowledge diffusion, Scientometrics 59(2) (2004) 199 211.

M. Chicurel, Bioinformatics: bringing it all together technology feature, Nature 419(6908) (2002) 751-757.

M. Coccia, The source and nature of general purpose technologies for supporting next K-waves: Global leadership and the case study of the U.S. Navy's Mobile User Objective System, Technol. Forecast. Soc. In Press (2016).

C.E. Cook, M.T. Bergman, R.D. Finn, G. Cochrane, E. Birney, R. Apweiler, The European Bioinformatics Institute in 2016: Data growth and integration, Nucleic Acids Res. 44(D1) (2016) D20 D26.

R.M. Cook-Deegan, The Gene Wars: Science, Politics, and the Human Genome, Norton, New York, NY, 1994.

N. Crafts, Steam as a general-purpose technology: a growth accounting perspective, Econ. J. 114(495) (2003) 338 351.

P.A. David, G. Wright, General purpose technologies and surges in productivity: historical reflections on the future of the ICT revolution, in: P.A. David, M. Thomas (Eds.), The economic future in historical perspective, Oxford University Press, (2006) 135 166.

D.B. Davison, D.T., Kingsbury, T.G. Marr, M.S. Waterman, Whither Computational Biology, J. Comput. Biol. 1(1) (1994) 1 2.

S. D'Souza, Gene Meets Machine: Intellectual Property Issues in Bioinformatics, Health L. Rev. 12(2) (2004) 34 43.

G. Dosi, Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change, Res. Policy 11(3) (1982) 147 162.

D. Eisinger, G. Tsatsaronis, M. Bundschus, U. Wieneke, M. Schroeder, Automated patent categorization and guided patent search using IPC as inspired by MeSH and PubMed, J. Biomed. Semantics 4(Suppl 1) (2013) 1 23.

P. Érdi, K. Makovi, Z. Somogyvári, K. Strandburg, J. Tobochnik, P. Volf, et al., Prediction of emerging technologies based on analysis of the US patent citation network, Scientometrics 95(1) (2013) 225 242.

M.P. Feldman, J.W. Yoon, An empirical test for general purpose technology: an examination of the Cohen-Boyer rDNA technology, Ind. Corp. Change 21(2) (2012) 249 275.

L. Fleming, O. Sorenson, Science as a map in technological search, Strategic Manage. J. 25(8-9) (2004) 909 928.

L. Fleming, O. Sorenson, Technology as a complex adaptive system: evidence from patent data, Res. Policy 30(7) (2001) 1019 1039.

L. Fleming, S. Mingo, D. Chen, Collaborative Brokerage, Generative Creativity and Creative Success, Admin. Sci. Quart. 52(3) (2007) 443 475.

D. Foray, P. David, B. Hall, Smart Specialisation – The Concept, Knowledge Economists Policy Brief 9 (2009) 1-5.

A. Gambardella, D. Harhoff, B. Verspagen, The Value of European Patents, Eur. Manage. Rev. 5(2) (2008) 69 84.

A. Gambardella, M.S. Giarratana, 2013. General technological capabilities, product market fragmentation, and markets for technology, Res. Policy 42(2) (2013) 315 325.

C. Goble, R. Stevens, State of the nation in data integration for bioinformatics, J. Biomed. Inform. 41(5) (2008) 687 693.

C.S. Greene, J. Tan, M. Unq, J.H. Moore, C. Chenq, Big data bioinformatics, J. Cell Physiol. 229(12) (2014) 1896 1900.

B.H. Hall, A.B. Jaffe, M. Trajtenberg, Market Value and Patent Citations, RAND J. Econ. 36(1) (2005) 16 38.

S. Hallam, Life out of sequence: a data-driven history of Bioinformatics, University of Chicago Press, Chicago, USA, 2013.

D. Harhoff, F. Narin, F., Scherer, K. Vopel, Citation Frequency and the Value of Patented Inventions, Rev. Econ. Stat. 81(3) (1999) 511 515.

R. Harrison, Protecting innovation in Bioinformatics and In-Silico biology, BioDrugs 17(4) (2003) 227 231.

E. Helpman, General purpose technologies and economic growth. MIT Press, Cambridge, MA, 1998.

M.O. Hill, Diversity and evenness: a unifying notation and its consequences, Ecology 54(2) (1973) 427 432.

D.W. Hosmer, S.A. Lemeshow, R.X. Sturdivant, Applied Logistic Regression. Wiley, Hoboken, NJ, 2013.

A. Jaffe, M. Trajtenberg, M.S. Fogarty, The Meaning of Patent Citations: Report on the NBER/Case-Western Reserve Survey of Patentees, NBER Working Paper No. W7631 (2000).

S. Kaplan, K. Vakili, The double-edge sword of recombination in breakthrough innovation, Strategic Manage. J. 36(10) (2015) 1435 1457.

G. Keijl, V.A. Gilsing, J. Knoben, G. Duysters, The two faces of inventions: the relationship between recombination and impact in pharmaceutical biotechnology, Research Pol. 45(5) (2016) 1061 1074.

M.A. Lemley, C. Shapiro, Probabilistic patents, J. Econ. Perspect. 19(2) (2005) 75 98.

J. Lewis, A. Bartlett, P. Atkinson, Hidden in the middle: culture, value and reward in Bioinformatics. Minerva 54(4) (2016) 471 490.

S. Li, T. Chambers, Y. Ding, G. Zhang, L. Meng, Patent citation analysis: calculating science linkage based on citing motivation, J. Assoc. Inf. Sci. Technol. 65(5) (2014) 1007 1017.

S. Lindmark, Coordinating the early commercialization of general purpose technologies: the case of mobile data communications, Innovation: Management, Policy & Practice 7(1) (2005) 39 60.

R.G. Lipsey, K.I. Carlaw, C.T. Bekar, Economic transformations: general purpose technologies and long-run economic growth, Oxford University Press, Oxford, 2005.

S.-T. Lo, D. Sutthiphisal, Crossover inventions and knowledge diffusion of general purpose technologies: evidence from the electrical technology, J. Econ. Hist. 70(3) (2010) 744 764.

S.K. Majumdar, O. Carare, H. Chang, Broadband adoption and firm productivity: evaluating the benefits of general purpose technology, Ind. Corp. Change 19(3) (2009) 641 674.

R. Margalef, Homage to Evelyn Hutchinson, or why there is an upper limit to diversity, Transactions of the Connecticut Academy of Arts and Sciences 44 (1972) 211 235.

A. Messeni Petruzzelli, D. Rotolo, V. Albino, Determinants of patent citations in biotechnology: an analysis of patent influence across the industrial and organizational boundaries, Technol. Forecast. Soc. 91 (2015) 208 221.

C. McCubbin, Legal issues in Bioinformatics, J. Commerc. Biotechnol. 9(3) (2003) 249 265.

D.G. McDonald, J. Dimmick, The Conceptualization and Measurement of Diversity, Commun. Res. 30(1) (2003) 60-79.

P. Murmann, K. Frenken, Toward a systematic framework for research on dominant designs, technological innovations, and industrial change, Res. Policy 35(7) (2006) 925 952.

A. Mushegian, Grand challenges in Bioinformatics and computational biology, Front. Genet. 2(60) (2011) 1 2.

Nature, Post-genomic cultures: Editorial, Nature 409(6820) (2001) 545.

H. Nakamura, S. Suzuki, I. Sakata, Y. Kajikawa, Knowledge combination modeling: the measurement of knowledge similarity between different technological domains, Technol. Forecast. Soc. 94 (2015) 187 201.

A.J. Nelson, Measuring knowledge spillovers: what patents, licenses and publications reveal about innovation diffusion, Res. Policy 26(3) (2009) 317 330.

E. Novelli, An examination of the antecedents and implications of patent scope, Res. Policy 44(2) (2015) 493 507.

M.V. Olsen, A time to sequence, Science 270(5235) (1995) 344 396.

T. Ostler, M. Gollin, Which types of bioinformatics inventions are eligible for patent protection? J. Commerc. Biotechnol. 21(2) (2015) 76 82.

C.A. Ouzonis, A. Valencia, Early Bioinformatics: the birth of a discipline – a personal view, Bioinformatics 19(17) (2003) 2176 2190.

H.S. Park, Preliminary Study of Bioinformatics Patents and Their Classifications Registered in the KIPRIS Database, Genomics Inform. 10(4) (2012) 271 274.

P. Patel, UK Performance in Biotechnology-related Innovation: An Analysis of Patent Data, Final Report prepared for the Assessment Unit of the UK Department of Trade and Industry, SPRU, (2003).

I. Petsas, The dynamic effects of general purpose technologies on schumpeterian growth. J. Evol. Econ. 13(5) (2003) 577 605.

E.C. Pielou, Species abundance distributions. In Ecological Diversity, Wiley Interscience, New York, NY, 1975.

M. Pop, S.L. Salzberg, Bioinformatics challenges of new sequencing technologies, Trends Genet. 24(3) (2008) 142 149.

J. Pustell, F.C. Kafatos, A convenient and adaptable package of DNA sequence analysis programs for microcomputers, Nucleic Acids Res. 10(1) (1982) 51 59.

B. Rasmussen, Innovation and Commercialisation in the Biopharmaceutical Industry: Creating and Capturing Value, Edward Elgar, 2010.

Research and Markets, Global Bioinformatics Market 2015-2019, (2015). Available at: http://www.researchandmarkets.com/reports/3244048/global-Bioinformatics-market-2015-2019#relb3 (Accessed on September 2015).

R.J. Robbins, Bioinformatics: essential infrastructure for global biology, J. Comput. Biol. 3 (1996) 465 478.

E. Sapsalis, B. van Pottelsberghe de la Potterie, R. Navon, Academic versus industry patenting: an in-depth analysis of what determines patent value, Res. Policy 35 (10) (2006) 1631 1645.

T. Savino, A. Messeni Petruzzelli, V. Albino, Search and recombination process to innovate: a review of the empirical evidence and a research agenda, International Journal of Management Reviews 19(1) (2015) 54 75.

P.P. Saviotti, M.A. DeLooze, S. Michelland, D. Catherine, The changing marketplace of Bioinformatics, Nat. Biotechnol. 18(12) (2000) 1247 1249.

A. Schaefer, D. Schiess, R. Wehrli, Long-term growth driven by a sequence of general purpose technologies, Econ. Model. 37(February) (2014) 23 31.

M. Schilling, E. Green, Recombinant search and breakthrough idea generation: an analysis of high impact papers in the social sciences, Res. Policy 40(10) (2011) 1321 1331.

L.I. Schultz, F.L. Joutz, Methods for identifying emerging general-purpose technologies: a case study of nanotechnologies, Scientometrics 85(1) (2010) 155 170.

D.K. Simonton, Creativity as blind variation and selective retention: is the creative process Darwinian? Psychol. Inq. 10(4) (1999) 309 328.

J. Singh, L. Fleming, Lone Inventors as Sources of Breakthroughs: Myth or Reality? Manage. Sci. 56(1) (2010) 41 56.

K.A. Smith, Laws, Leaders, and Legends of the Modern National Library of Medicine, J. Med. Libr. Assoc. 96(2) (2008) 121 133.

R. Staden, Sequence data handling by computer, Nucleic Acids Res. 4(11) (1977) 4037 4051.

A. Stirling, On the economics and analysis of diversity, SPRU Electronic Working Paper No. 28, (1998). Available at: http://www.microtheory.uni-jena.de/download/Stirling1998.pdf (Accessed on January 2016).

Strategic Direction International (SDI) Inc., Bioinformatics: The Rosetta Stone of Life Science Research, Instrument Business Outlook, (2001).

C. Taylor, Bioinformatics: Jobs Galore, Science Careers, (2000). Available at: http://sciencecareers.sciencemag.org/career_magazine/previous_issues/articles/2000_09_01/nodoi.8547776132534311194 (Accessed on November 2015).

G. Thoma, Striving for a large market: evidence from a general-purpose technology in action, Ind. Corp. Change 18(1) (2009) 107 138.

M. Trajtenberg, A Penny for Your Quotes: Patent Citations and the Value of Innovations, RAND J. Econ. 21(1) (1990) 172 187.

M. Trajtenberg, A. Jaffe, R. Henderson, University versus Corporate Patents: A Window on the Basicness of Invention, Economics of Innovation and New Technology 5(1) (1997) 19 50.

B. Uzzi, S. Mukherjee, M. Stringer, B. Jones, Atypical combinations and scientific impact. Science 342(6157) (2013) 468 472.

A. Van Zon, T. Kronenberg, General purpose technology and energy policy, in: L. Bretschger, S. Smulders (Eds.), Sustainable resource use and economic dynamics, Netherlands, Springer, (2007) 71 99.

C. Vorndran, R.L. Florence, Bioinformatics: patenting the bridge between information technology and the life science, IDEA: The Journal of Law and Technology 42(1) (2002) 93 132.

J. Youtie, M. Iacopetta, S. Graham, Assessing the nature of nanotechnology: can we uncover an emerging general-purpose technology? J. Technol. Transfer 33(3) (2007) 315 329.

M.S. Waterman, Introduction to computational biology: maps, sequences and genomes, Chapman & Hall, 1995.

T.J. Welch, Bioinformatics: the role and limitations of patents, Drug Discov. Today 7(11) (2002) S77 S82.