Regular article

Measuring the knowledge translation and convergence in pharmaceutical innovation by funding-science-technology-innovation linkages analysis

Jian Du*, Peixin Li, Qianying Guo, Xiaoli Tang

Institute of Medical Information & Library, Chinese Academy of Medical Sciences, Beijing, 100005, China

ARTICLE INFO

ABSTRACT

We propose a backward tracking model for measuring knowledge transfer in the whole translational research spectrum. Using the drugs-patents-papers-grants backward linkages, we try to figure out the funding-science-technology-innovation translational pattern and ponder some policy implications on e.g., which priority areas and knowledge convergence level are more likely to generate new drugs. The drug-patent linkage data was accessed through the USFDA Orange Book, covering a drug's active ingredient, formulation, or methods of use for approved indications. It will take about 10 years from the application of earliest patent to the approval of the new drug. Also such high-value patents in FDA Orange Book tend to cite scientific knowledge published on average 10–15 years ago. The technology linkage of new drugs was relatively stable while the science linkage of technology inventions increased rapidly. Among the scientific papers cited by drug patents, private-institution originated papers are only a quarter of the public. By linking theses scientific papers with funding sources, we found a large majority (90%) are public-funded and only a very small part are private-funded or public-private joint-funded. Our study also indicates the importance of research on such fields as pharmacology, chemistry (including medicinal chemistry, biochemistry, and organic chemistry), molecular biology, neurosciences, and immunology on new drugs innovation. There is no obvious relationship between "basicness" and linkages to the resulting patents' impact and to drugs innovation. A balanced basic research and applied research maybe essential for fostering drug innovation because it is a complete chain translating from basic discovery to clinical evidence then to clinical practice. In order to achieve successful pharmaceutical innovation, rather than focusing on only technology, convergence with science at moderate levels (maybe 1/3) is suggested.

© 2018 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Biomedical innovation is the process of transforming scientific discoveries into vaccines, diagnostics, devices and therapies that improve and save patients' lives, which is a typical process of translational research. The pace of basic discoveries in biomedical research is accelerating; however, translation of this knowledge into concrete improvements in clinical medicine

continues to lag behind the pace of discovery (Duda et al., 2014). Estimated by both life expectancy and New Molecular Entities approved by the United States Food and Drug Administration (Bowen & Casadevall, 2015), showed that biomedical research outcomes over the last five decades have remained relatively constant despite rising resource inputs and scientific knowledge. They reminded the crisis of the growing input-outcome disparity in biomedical research, which emphasized the importance of translational research. If policy interventions are going to be designed and implemented, policy makers need to know where translational research is happening, how long the translation time lag is, and how to reach successful translation.

Previous studies for tracking and evaluating translational research have focused on citation analysis at the article level. For example, by investigating the citation relationship between different kinds of research papers, namely clinical observation, clinical mix, clinical research, and basic research, the knowledge translation between basic research and clinical research could be measured and translational interface and translational paths could be revealed (Cambrosio, Keating, Mercier, Lewison, & Mogoutov, 2006; Jones, Cambrosio, & Mogoutov, 2011). By probing into the papers cited by clinical guidelines and public health guidelines, one can measure the impact of basic research on clinical practice (Grant, Cottrell, Fawcett, & Cluzeau, 2000; Thonon, Saghatchian, Boulkedid, & Alberti, 2015). In addition to the linear measurement between the two points or phases in translational research (Weber, 2013), developed a bibliometric approach to map PubMed articles onto a graph, called the *Triangle of Biomedicine*. The corners of the triangle represent research related to animals, cells and molecules, and humans; and, the position of a publication on the graph is based on its topics, as determined by its Medical Subject Headings (MeSH). He defined translation as movement of a collection of articles, or the articles that cite those articles, towards the human corner. Then the Office of Portfolio Analysis of National Institute of Health (NIH) further developed iTrans, a powerful web application mapping published articles along three axes: cellular/molecular research, animal research, and human research. iTrans can be used for measuring and visualizing translational status of research over time, and even detecting early signs that of bench-to-beside translation is starting (National Institute of Health, O. o. P. A., 2014). The common notion found in these studies is that citation network analysis among publications with different research levels in the translational research spectrum can be used for measuring and identifying the emergence and interface of translational research.

The above studies focused on the forward citations at the article level. Previous literature has also used science related references from non-patent literature (NPL) as a tool to represent the direct relationship between an technology and scientific knowledge since the pioneering work by Narin and colleagues (Carpenter, Cooper, & Narin, 1980). The non-patent literature cited by a patent document measures the translation from basic research to technology invention (Narin & Noma, 1985), while patent documents cited by scientific articles reveal the penetration of technology innovation into basic research (Glänzel & Meyer, 2003). Several studies have suggested that the biomedical patents cited more basic research publications and thus the biomedical industry relied on public science more than other industries (Jibu, 2014; McMillan, Narin, & Deeds, 2000; Sung, Wang, Huang, & Chen, 2015). Measuring citations to scholarly works in the global patent literature enables assessment of the influence of published research on invention, industry and enterprise, at the individual and institutional level. While these metrics based on the forward citations at the paper level or paper-patent level can be used to measure translational research, they lack precision and fail to cover the entire translational spectrum.

Most recently, a few studies extended the entire translational spectrum from grants to papers, then to patents and ultimately to drugs. But the aim of such studies is to investigate the applied value of public funding in biomedical research instead of measuring the pattern of translational biomedical research. For instance, using data on patents linked to U.S. National Institutes of Health (NIH) grants over a 27-year period (Li, Azoulay, & Sampat, 2017), provide a large-scale accounting of linkages between public research investments and subsequent patenting. The results show about a third of NIH grants generate research that is cited by commercial patents. This is much greater than the share of grants directly yielding patents (less than 10%), even though policy-makers often focus on this easier-to-grasp metric to capture the near-term economic returns to public funding of biomedical R&D. (Cleary, Beierlein, Khanuja, McNamee, & Ledley, 2018) examined the contribution of NIH funding to published basic research on biological targets as well as applied research associated with new molecular entities (NMEs) approved by the Food and Drug Administration. They suggested that the NIH contribution to research associated with new drug approvals is greater than previously appreciated and highlighted the risk of reducing federal funding for basic biomedical research. By linking public funding with the resulting knowledge and medicines, we can observe the impact of such funding which was translated into socially valuable outcomes. Such linkages are easier to assert than to prove.

So, in this paper we try to develop a relative precise quantitative framework for measuring knowledge transfer in the whole translational spectrum through direct linkages between research grants and papers and patents and drugs. We will provide empirical data for the linkages between science (publications), technology (patents) and innovation (commercialized products, such as new drugs). The framework reflects not just the correlation between several artefacts, instead, it reveals as close as possible the causality between inputs and activities and subsequent outputs and tangible outcomes. In this study we provide evidence of (i) the translational delay, namely the time lag between the scientific knowledge of a potential drug and its market approval, (ii) science linkage and the technology linkage of the innovative drugs, (iii) scientific focus and structure acting as a source of inspiration in technical inventions of new drugs, (iv) to what extent the convergence of scientific and technical knowledge impact on successful pharmaceutical innovation, and (v) structure of research funding for scientific publications cited by the drug patents. Using the drugs-patents-papers-grants backward linkages, we try to figure out the
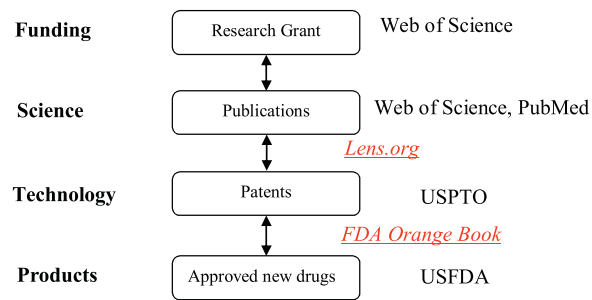
Funding	Research Grant	Web of Science
	↕	
Science	Publications	Web of Science, PubMed
	Lens.org	
	↕	
Technology	Patents	USPTO
	FDA Orange Book	
	↕	
Products	Approved new drugs	USFDA

Fig. 1. The 'backward tracking model' for measuring translational research.

funding-science-technology-innovation translational pattern and ponder some policy implications on e.g., which priority areas and knowledge convergence level are more likely to generate new drugs.

2. Methods and data

2.1. Conceptual framework

We propose a backward tracking model for measuring knowledge transfer in the whole translational research spectrum. It is necessary to match the measureable linkages between various artefacts of the translation process. We focused on the backward tracking since the data of drugs approved by the United States Food and Drug Administration (USFDA) was set as the milestone of translational terminal. The data of key patents referring to FDA approved drugs was accessed through the FDA Orange Book, in which the pharmaceutical companies listing the key patents that protect their drugs when applying the FDA's new drug approval. The data of papers was retrieved from the non-patent references cited by the drug patents. The research grant information has been recorded in the acknowledge section of a publication, we use these data to link every grant, whether government granted or industry granted, with the papers that result. Based on the approved drugs, the key patents can be backward tracked. Then the scientific papers cited by the key patents can be further captured, and then the funding information of these papers can also be extracted. By a series of backward tracking, the translational process among grants-papers-patents-drugs can be revealed (Fig. 1).

2.2. The drug-patent linkage

Why using USFDA approved new drugs as the original data? The United States is the world's largest pharmaceutical market, comprising roughly 40% of the world's pharmaceutical revenues and, thus, always acts as benchmark in the scientific drug regulatory system. Moreover, since its relatively rigorous regulatory system of drugs and their intellectual property, the US is always chosen as the first target market by the majority of worldwide pharmaceutical firms, making it a good data source to evaluate pharmaceutical innovation (Friedman, 2010, 2017). An additional benefit to focusing on drugs marketed in the US stems from its drug-patents linkage system. Since the 1984 Hatch-Waxman Act, branded firms are required to list on the Orange Book (officially named Approved Drug Product with Therapeutic Equivalence Evaluations) patents issued before drug approval with at least one claim covering a drug's active ingredient, formulation, or methods of use for approved indications. In recent years, the drug-patents linkages originated from FDA Orange Book have been used to investigate the geographical distribution of pharmaceutical innovation, and to measure the technology capacity and patents strategy of pharmaceutical firms (Friedman, 2010, 2017; Hu, Scherngell, Man, & Wang, 2013; Huang, Fang, & Chang, 2011). Linkages between drug and patents provide objective, non-manual annotated, and thus unbiased data for identifying the products' key technologies.

2.3. The patent-paper linkage

In general, patent references can be divided into backward citation references (cited by the focal invention) and forward citation references (other sources citing the focal invention). Focusing on the backward citation references enables identification of prior knowledge which inspires the invention process, while analyzing the forward citation references allows for tracing descendant knowledge such as inventions which were influenced by the patent. Backward citation references are further divided into patent references and non-patent literature (NPL) which consists of references to journal articles, conference proceedings, books, databases, textbooks, corporate reports and other documents. Earlier patents may be cited by the inventor to demonstrate their difference from prior art or added by the examiner to limit the scope of the patent. Patent backward citations to NPL are considered stronger indicators of the impact of economically-valuable scientific knowledge on business patenting than citations to patents (Roach & Cohen, 2013). To consider the influence of science related references, this study limited the scientific knowledge to journal articles which were published in Web of Science and/or PubMed listed

journals only. So, using the patent backward citations to journal papers, one can measure knowledge flow from scientific research to technical inventions, and so this approach has been applied for this analysis.

2.4. The paper-grant linkage

We want to know which type of funding agencies generates more research cited by patents and subsequently lead to new drugs approval. The focus task is to capture the research funding agencies for each article which has been cited by drug patents. In recent years, by using Web of Science (WoS) funding acknowledgement fields as a standard and publicly available dataset linking research publication outputs with funding sources, funding acknowledgement analysis has been assessed as an enhanced tool to investigate research sponsorship impacts (Wang & Shapira, 2011). Unfortunately, the WoS started in August 2008 to systematically include funding acknowledgement data in its publication records. Other well-known public databases, such as Scopus and PubMed also have not completely include research grant data in its publication records, and thus could not be used in this study. According to a recent investigation, for WoS databases only funding acknowledgments in journals indexed by SCI are systematically recorded. In other words, the WoS funding acknowledgment data is not suitable for analyzing social science and humanities research (Tang, Hu, & Liu, 2017). Nevertheless, funding acknowledgments analysis opens a wide range of possibilities for linking scientific input and output.

2.5. The data sources of the drug-patent-paper-grant linkages

In the present study, data from the Orange Book about the approved new drugs was collected from the USFDA *Approved Drug Products with Therapeutic Equivalence Evaluations* website through Orange Book Data Files (updated monthly)[1]. The Orange Book primarily contains three parts: 'Product', 'Patent', and 'Exclusivity'. We capture the US patent number and drug application number from the file 'Patent', and match the drug application number from the file 'Product' to the corresponding drug information, such as Drug approved date, Ingredient, Trade Name and so on. Drugs approved between 2006 and 2015 were included in this study, and 605 drugs and the corresponding 2422 USPTO patents were identified. FDA classifies drugs according to the novelty of their chemical substance (i.e. new chemical or molecular entities (NMEs) vs. updates, i.e. substances that already have been under review by the authority) and therapeutic potential which determines review speed (priority review for drugs fulfilling a high medical need vs. standard review). So the 605 drugs include both the NMEs and the updates of substances that already have been under review by the authority. We collect patent data from USPTO Patent Full-Text and Image Database (PatFT) to capture the relevant information (e.g., assignee name, application date, publication date, and nationality of inventors).

In a patent submitted to USPTO, the assignee and the examiner should list references to the sources of knowledge which were used in the invention process. We capture the scholarly articles cited by these 2422 patents using US patent numbers from Lens.org (https://www.lens.org/lens/patcite) to explore the cited scholarly work found in patent literature. It is a public benefit platform developed by Cambia (a global non-profit social enterprise) and the Queensland University of Technology in Australia. The Lens.org, including its predecessor called Patent Lens, have been a free, open and global full text patent search site for over 18 years, serving over 110 M patent records from more than 100 jurisdictions. The methodology of mapping scholar papers with non-patent literature in patents is well articulated in the two recent articles published in *Nature* (Jefferson, 2017) and *Nature Biotechnology* (Jefferson et al., 2018). The platform has linked the world's patent information to most of the scholarly literature using DOI and PMID by collaborating with CrossRef and National Library of Medicine. Using data from Les.org, the Nature Index 2017 Innovation[2] ranked top 200 institutions whose academic research having impact on innovation by examining how research articles are cited in third party patents. After extracting scientific papers cited by patents via lens.org, our dataset involved 14,831 papers with PMID and 14,228 papers with DOI. After removing duplicates, 15,055 papers in the collection of Web of Science provided by Clarivate Analytics were identified.

The funding information of these papers was accessed via Web of Science funding acknowledgement fields, which involved 134 funding agencies/organizations. Note that Web of Science began systematically collecting acknowledgment information since 2008 in SCI-Expanded and since 2015 in SSCI. So, the data is not completed for interpretation and we will discuss the detail in the following section.

3. Results

3.1. General information

Based on a series of backward tracking, our dataset includes 605 drugs, citing 2422 patents, citing 15055 SCI-indexed papers, acknowledging more than 130 funding agencies. Among the 15,055 scientific papers cited by new drug patents, 41% were published in the 1990s, 35% were published in the 2000s, and 14% in the 1980s. Research papers published during the
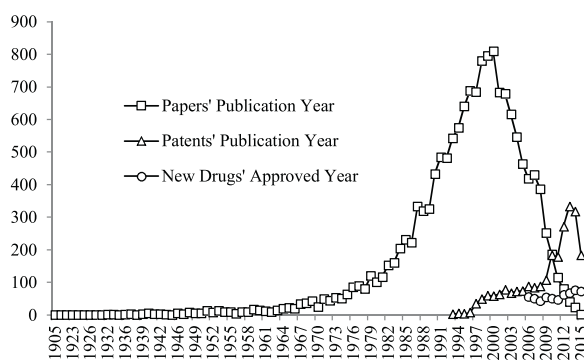

Fig. 2. Papers' publication year, patents' publication year and new drugs' approved year.

1970s and 2010s accounted for 3%–4%, respectively. In general, two thirds (76%) papers were published between 1990 and 2010 (Fig. 2).

The earliest paper with the title "Information on the five isomeric acids C4H6O2", was published at 1905 in a German-language scientific journal *Berichte der deutschen chemischen Gesellschaft* (now *European Journal of Inorganic Chemistry*), which was founded in 1868 and has been one of the oldest scientific journals in chemistry. This paper was cited by the US patent (US7491725), filed in July 29, 2005 with the title "Process for preparing 2-aminothiazole-5-aromatic carboxamides as kinase inhibitors". The patent was owned by Bristol-Myers Squibb Company as the assignee. This paper was cited by the patent in this such context as "*electrophilic bromination of acrylanilide and crotonanilide has been known to undergo both aromatic bromination and addition to the. alpha.,,beta.-unsaturated carbon-carbon double bonds.*" The terms "*. . . has been known . . .*" reflected the linkage between the scientific knowledge paper and technical invention patent.

Then the US7491725 patent was added by Bristol-Myers Squibb Company to the FDA orange book to protect the intellectual property of a new drug with Trade Name "SPRYCEL" and "DASATINIB" as the Active Ingredient. It was approved to market by USFDA in Jun 28, 2006 with Strength (20 MG/50MG/70 MG) and in May 30, 2008 with Strength (100 MG) and in Oct 28, 2010 with Strength (140 MG).

3.2. The time lag of translation

In pharmaceuticals, the path from basic research towards marketable drugs is standardized, starting with the synthesis of the compound, various clinical studies representing feedback loops for the actual development process, and – last but not least – market approval. The path is long-lasting. According to (Sternitzke, 2010)'s overview about the empirical studies, it will take on average 8–14 years between compound synthesis and commercialization such as market approval. In order to reveal the time lag of translation, we compared the application year of the earliest patent in the FDA Orange Book and the approval year by FDA for each drug. In general, patents last 20 years. After expiration, the drugs will become OTC (over the counter). In order to extend the patents, the assignees usually make some so called enhancement and refiled for new patents in the same drug. This will make them keep the patents for a longer time. That's why we can see two or three patents were listed for the same or similar things. This focus on early patents was used to avoid including research into derivatives or enhanced formulations; the follow-on patents may be materially different from the initial patents. In general, the time lag increased from 7.6 years in 2006 to 10.3 years in 2015 (Fig. 3). In other words, nowadays it will take about 10 years from the first application of key patent to the approval of the new drug.
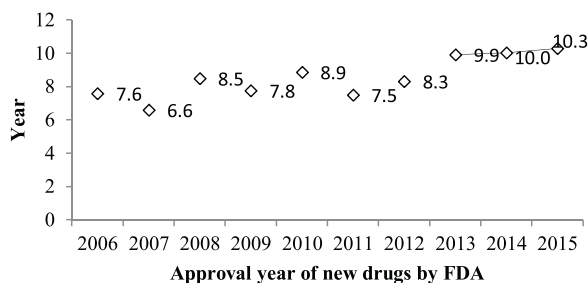

Fig. 3. The time lag of translation between patent applications and new drugs approval.
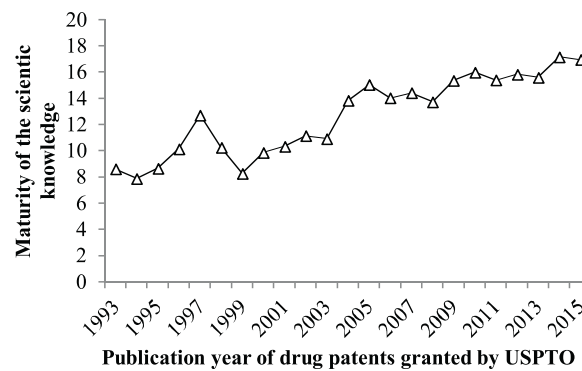
Fig. 4. The time lag between the average publication year of the non-patent references and the granted year for each drug-patent.

3.3. The extent of knowledge maturity used for the technological inventions

When concerning the time lag between the basic science paper and the patent in the FDA Orange Book, we turned to another perspective because it is more difficult to identify drug's direct scientific knowledge than technical knowledge. The non-patent literature cited by the drug patent is not always exactly the right and direct basic science paper that contributed to the drug's development. By manually defining the most recent basic science papers as leading to understanding or description of a biologic process without studying any drug candidate for the 15 drugs approved by FDA in 2011, (Reidenberg & Erle, 2012) found that the earliest patent in the FDA Orange Book for each of averaged 2 years after the most recent basic paper. So, it seems that there is almost no big delay between the most relevant basic paper and the earliest patent directly leading to the innovation, since the granting of a patent also involves assigning intellectual property based upon an original invention or discovery to establish "priority of discovery". The primary goal of a patent is to stake territory as one's own and to exclude others (or make them pay for using your intellectual property), thereby serving the goal of commercialization (Vale & Hyman, 2016).

So, this paper will not prepare to calculate the time lag between the scientific knowledge and patented technology. Instead, we introduce the notion of *knowledge maturity*. It is defined as "the time elapsed between the original discovery of that knowledge and its incorporation in a new invention" (Capaldo, Lavie, & Messeni Petruzzelli, 2017). Cutting-edge knowledge-based invention usually suffers from limited ways of applications as well as requires additional tests to prove it. As time goes by, inventions based on matured knowledge are shown to be more reliable and applicable because sufficiently matured knowledge is investigated in-depth and has proven its usefulness. In addition, matured knowledge becomes codified and thus can be more easily transferred and understood between researchers. In this notion, the maturity of scientific knowledge determines the efficiency of knowledge searching in convergence. Similar with (Lee, Park, & Kang, 2018), we calculate the time lag between the average publication year of the non-patent references and the granted year for each patent, and identify the extent of knowledge sources' maturity for each technology. A similar increasing trend was also found for the maturity of the scientific knowledge used by the drug patent (Fig. 4). Because all new knowledge is tentative and subject to revision, it can take a period of time and contributions by many researchers until the initial conjecture comes to be regarded as a "discovery" by the community. On average, in the more recent years, patents increasingly use matured and thus tested and proven scientific knowledge, concepts or theories. The maturity of the scientific knowledge has increased from 10 years for patents granted between 1993 and 2003 to 15 years for patents granted during 2004–2015.

3.4. Science linkage and technology linkage

The linkage between science and technology has obtained the interest of many researchers in the fields of bibliometrics and research policy. The Science Linkage (SL) of a technology was defined by the CHI Research as the number of scientific papers cited by a patent, to measure the degree of linkage between scientific research and technological inventions. Among the 2422 patents in this study, 1972 of them, accounting for 81.4% in their backward references cited scientific papers. Such a high proportion indicates that drug patents rely heavily on knowledge derived from scientific research. Taking into account the paper-patent-product linkage, we introduce a new indictor, namely the Technology Linkage (TL) of an innovation, as the number of key patents listed in the FDA Orange Book to protect the intellectual property of the drug product. As is shown in Fig. 5, the science linkage of technological inventions in pharmaceutical industry shows a remarkable growth over time. The number of scientific papers cited per patent increased from 5 in the 1990s to nearly 30 in the 2010s. By contrast, the technology linkage of drug products keeps stable. On average four or five patents were needed to protect the intellectual property of a new drug in the applicant's patent portfolio. If a drug finally has been developed, firms apply different strategies for the further development and protection of pharmaceutical innovations in order to prevent imitation by competitors such as generic drug makers. These practices are known as drug life cycle management or 'evergreening'. A number of consecutive patents for new combinations, uses, formulations, manufacturing processes, or molecules follow the original or basic patent.
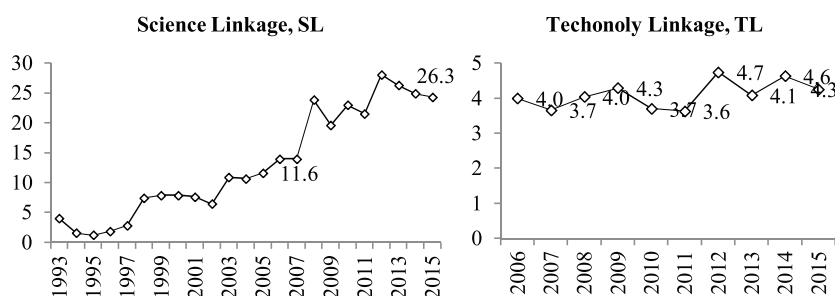
Fig. 5. Science Linkage of technological inventions and Technology Linkage of drug products.

Table 1
Top 10 subject categories and published journals of the papers cited by drug patents.

Subject Categories	N	%	Published Journals	N	%
Pharmacology Pharmacy	4639	30.8	*Journal of Medicinal Chemistry*	466	3.1
Chemistry	1984	13.2	*Journal of Pharmaceutical Sciences*	224	1.5
Biochemistry Molecular Biology	1462	9.7	*International Journal of Pharmaceutics*	218	1.4
Neurosciences Neurology	992	6.6	*Proceedings of the National Academy of Sciences of the United States of America*	210	1.4
General Internal Medicine	991	6.6	*Journal of Biological Chemistry*	199	1.3
Oncology	866	5.8	*New England Journal of Medicine*	195	1.3
Endocrinology Metabolism	652	4.3	*Journal of Organic Chemistry*	180	1.2
Science Technology Other Topics	586	3.9	*Pharmaceutical Research*	179	1.2
Immunology	584	3.9	*Antimicrobial Agents and Chemotherapy*	176	1.2
Cell Biology	464	3.1	*Cancer Research*	157	1.0

Combining science linkage and technology linkage, it can be seen that the key technologies of drug products increasingly origin from scientific research.

3.5. Scientific publications cited by the drug patents

Table 1 displays the top 10 most cited research fields defined by Web of Science subject categories. Almost one third of the 15,055 scientific papers cited by drug patents are from the field of pharmacology and pharmacy. In addition, within life sciences, chemistry make up the largest share (13%), followed by biochemistry and molecular biology (24%), immunology (4%), and cell biology (3%). In clinical sciences, neurosciences and neurology contribute the largest share (6.6%), followed by general and internal medicine (6.6%), oncology (5.8%), and endocrinology and metabolism (4.3%). The distribution of journals and research field, indicated the importance of research on such fields as pharmacology, chemistry (including medicinal chemistry, biochemistry, and organic chemistry), molecular biology, neurosciences, and immunology on new drug discovery and development.

Table 2 shows the distribution of authors' nations/regions of 13,918 papers (published after 1980s, account for 92.4% of all 15,055 papers) which have been included in Incites from Clarivate Analytics. US contributes the largest proportion (53%) of publications which cited by the drug patents around the world. UK accounts for 12%, Germany 7.8% and Japan 6.7%. The Category Normalized Citation Impact (CNCI) of the top 30 countries/regions' publications are all larger than 1, indicating that publications cited by drug patents tend to show a higher citation impact. An industry collaborative publication is one that lists its organization type as "corporate" for one or more of the co-author's affiliations. The % of Industry Collaborations is the number of industry collaborative publications for a country divided by the total number of documents for the same country represented as a percentage. Another pattern of these papers is also noticed that publications cited by drug patents at a large proportion originate from such studies carried out by the academy collaborated with the industry.

3.6. The structure of scientific knowledge

Next, we will investigate the structure of scientific knowledge to answer the questions such as whether they are more originated from "public" or "private" institutions and to what extent they are more "basic" or "applied"

### 1) *"public" or "private"?*

The non-patent references published between 1905 and 2015 in our dataset are based on Web of Science database. Of all the SCI-indexed papers, 82% have the information of author address and 76% (11,609) have the information of corresponding author's address. This percentage is in accordance with a recent investigation by (Liu, Hu, & Tang, 2018) which revealed that

Table 2
Distribution of authors' nations/regions of scientific papers cited by drug patents.

	nations/regions	N	%	CNCI	% collaborated with industry
1	US	7382	53.0	5.8	8.9
2	UK	1678	12.1	5.9	14.1
3	Germany	1090	7.8	4.5	10.9
4	Japan	934	6.7	3.8	3.3
5	France	704	5.1	6.1	11.8
6	Canada	668	4.8	6.0	10.6
7	Italy	633	4.5	3.6	7.6
8	Sweden	453	3.3	4.2	13.9
9	Netherlands	370	2.7	5.2	10.3
10	Switzerland	349	2.5	4.9	22.4
11	Australia	341	2.5	5.8	7.9
12	Denmark	278	2.0	5.6	16.6
13	Belgium	268	1.9	5.9	23.1
14	Spain	261	1.9	3.8	9.2
15	Israel	188	1.4	5.1	10.6
16	Finland	136	1.0	2.6	5.9
17	Austria	120	0.9	4.2	14.2
18	New Zealand	118	0.8	5.5	7.6
19	India	99	0.7	1.9	2.0
20	Norway	80	0.6	6.4	16.3
21	Poland	77	0.6	6.1	20.8
22	Korea	68	0.5	1.8	5.9
23	China Mainland	59	0.4	2.3	8.5
24	Brazil	51	0.4	3.7	5.9
25	Portugal	51	0.4	8.3	17.7
26	Taiwan	50	0.4	1.9	12.0
27	Greece	50	0.4	4.1	12.0
28	Ireland	47	0.3	4.7	10.6
29	Turkey	46	0.3	4.0	6.5
30	Argentina	41	0.3	4.9	19.5

from 1900 to 2015 over one-fifth of the publications in the three core journal citation databases of Web of Science have completely missing information from the address field. The principal investigators (PI) are scientists who play a leading role in an independent research team. They (usually senior advisors) are responsible for the direction of the research team, design the roadmap of a project, assign research tasks, guide the experiments, and check the logic of the resulting articles. Therefore, they are among the most important people in the research team and usually act as the corresponding author in multi-authored articles. While they usually do not take part in the experiments, someone is assigned by them to play the core role in each project and write a paper as the first author. The other partners with minor contributions to the research project usually sign as co-authors in the resulting article. So, we use the information from the corresponding author's address field to represent the most important contributions. The origin of the 11,609 non-patent references cited by these drug patents was assigned to two categories. The first category, the public sector, comprises all universities, medical schools, private non-profit research institutes (such as Howard Hughes Medical Institute), and hospitals, but also governmental labs/institutions like the National Institute of Health or the National Cancer Institute. Private firms constitute the second category. We checked the origin of the corresponding address for each article, and classified them into Public and Private. The results show that there are 8964 Public papers, about four times that of the Private (2209 papers), as is shown in Fig. 6.
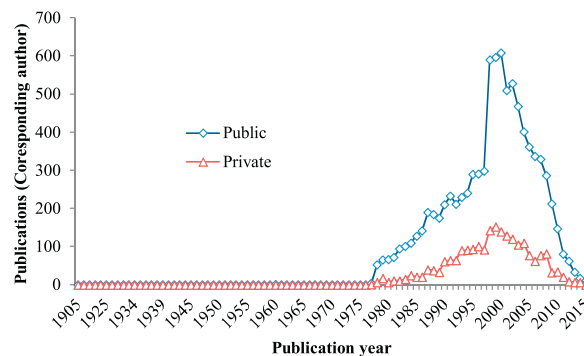
Fig. 6. Institutional source of cited articles by drug patents.

2) *"basic" or "applied"?*

The question of whether more "basic" or "applied" research ultimately more valuable for successful drug innovation is of great interest in science policy. One complication is that there is no consensus on the definitions and distinctions between the two. "Basic" research has been variously defined by whether it seeks general or specific knowledge, by the institutional environment where it takes place and the norms regarding dissemination, by whether it is undertaken for its own sake or with some application in mind, and by whether or not it is targeted to a specific program or mission, among other ways.

The 'basicness' of the articles was usually assessed by means of the CHI Research (now ipIQ) pioneering work on the structure of biomedical literature (Narin, Pinski, & Gee, 1976). The classification relied on the type of journal where the article is published. The four-levels of the classification range from clinical observation (level 1) to clinical mix, clinical investigation, and basic research (level 4). (Lewison & Paraje, 2004) updated the system and proposed a new method of classification of biomedical research journals by research level into clinical or basic, or somewhere in between. It is based on counting articles that have one of about 100 "clinical" title words, or one of a similar number of "basic" title words, or both. It allows over 3000 journals in the Science Citation Index (or other databases) to be classified rapidly and transparently, for changes in their research level with time, and for many individual papers in "mixed" journals to be categorised as clinical or basic. However, currently less than half of all articles published are in journals that been classified by research level (Boyack, Patek, Ungar, Yoon, & Klavans, 2014).

Rather than try to resolve this debate, we examine three different dimensions that have been of interest to medical research policy-makers: whether the research is disease-oriented, whether it is focused on patients, and whether the original publications clearly state that the study might have future clinical therapeutic or preventive application. Except for the last measure, the former two "basicness" measures rely on Medical Subject Heading (MeSH) terms, the controlled vocabulary maintained by the National Library of Medicine. The NLM Medical Subject Headings controlled vocabulary of biomedical terms that is used to describe the subject of each journal article in MEDLINE. MeSH contains approximately 26 thousand terms and is updated annually to reflect changes in medicine and medical terminology. We use 14,831 PubMed publications to conduct the following analysis.

A paper is said to be disease-oriented if it has at least one MeSH Major Topic term corresponding to a disease (i.e., the MeSH code starts with the letter C). MeSH Major Topic term is defined as a MeSH term that is one of the main topics discussed in the article denoted by an asterisk on the MeSH term or MeSH/Subheading combination, e.g., Cytokines/physiology*. By this measure, 7148 papers (48% of our sample) are disease-oriented. Distinguishing patient-oriented papers from other papers is straightforward, because the MeSH controlled vocabulary includes a term for humans. Patient-oriented papers defined in this way include (but are not limited to) research that uses human subjects. Using this measure, 10,155 grants (68% of our sample) are patient-oriented. We identified therapy targeted papers in all 14,831 PubMed publications that their title and/or abstract fields contained the word therapy, therapies, therapeutic, therapeutical, prevention, preventive, vaccine, vaccines, or clinical. These articles clearly stated that the study might have future clinical therapeutic or preventive application. By this measure, 5833 grants (40% of our sample) are therapy-oriented.

We regard diseases targeted papers as applied research and not diseases targeted papers as basic research. Similarly, we understand human targeted papers as applied research and not human targeted papers as basic research. Also, we consider therapy targeted papers as applied research and not therapy targeted papers as basic research. We investigate are there any differences for each pair of paper groups when observing the scientific knowledge contribution to new drugs. One is the share of papers in each group. The other is forward citations per patent, indicating the technical impact of the resulting patents. Fig. 7 shows a balanced proportion, namely each accounted for 50% on average, between basic or applied oriented research contributing to new drugs innovation. Papers with disease and human orientation result patents with a similar impact reflected by forward citations per patent. Non–human-oriented research yields patents at virtually identical average impact to human-oriented research. The orange line and blue line almost coincide. Taken together, Fig. 8 suggest that, based on our measures, basic and applied research are quite similar in their linkages to the resulting patents' impact and to drugs innovation.

3.7. Science-technology interactions for new drugs

The impact of discoveries may extend beyond the domain of science and may be crucial steps towards technological applications, and to innovations and products. Scientific discoveries and their incorporation in technology are often inter-linked in complex ways within research and development (R&D) systems. To reveal as much as possible the whole picture of research areas for the scientific papers and the interactions with the technical focus of drug patents, we firstly match the WoS-journals with Essential Science Indicators (ESI) Research Areas[3]. In ESI, a journal can be assigned to only one field. Journals such as *Science* and *Nature* are categorized as multidisciplinary since they publish research in many different fields. 13,551 papers (account for 90%) were match to one of the ESI Research Areas. One third of scientific knowledge is from clinical medicine, followed by pharmacology (about a quarter) and Chemistry (including biochemistry, accounting for 20%).

---

[3] http://help.prod-incites.com/inCites2Live/filterValuesGroup/researchAreaSchema/esiDetail.html.
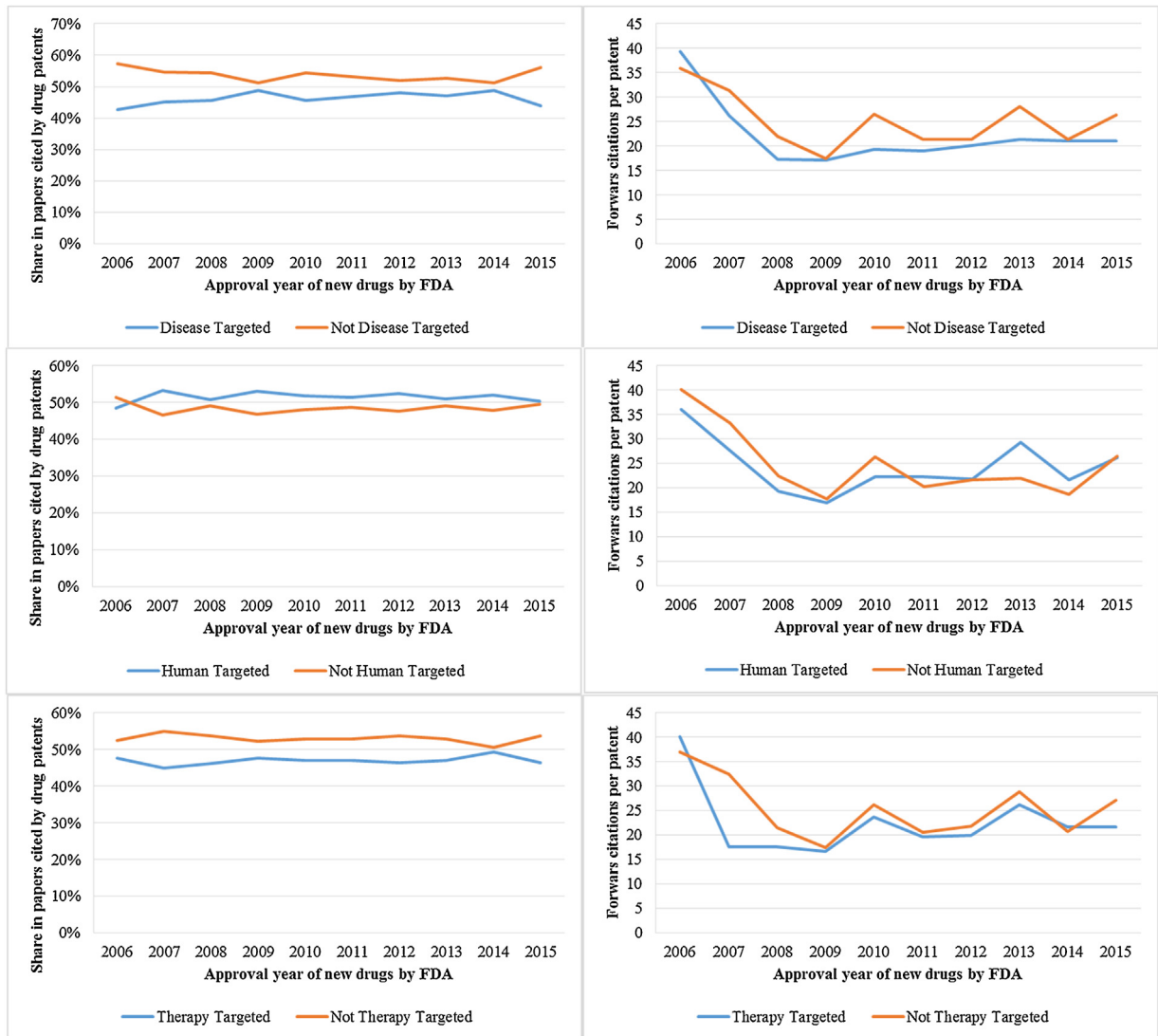
Fig. 7. The contribution share to drug innovations and the average impact of the resulting patents for basic or applied oriented research.
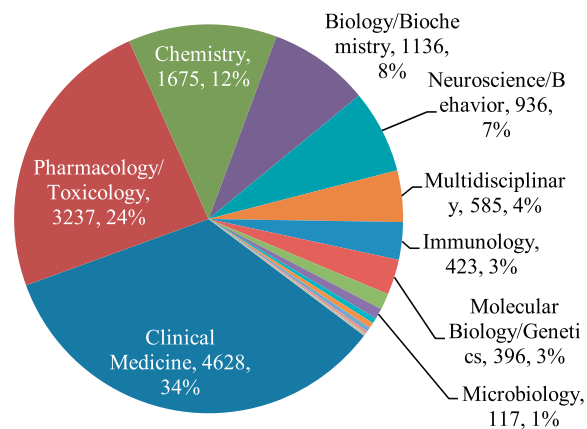

Fig. 8. The distribution of scientific papers by Essential Science Indicators Research Areas.
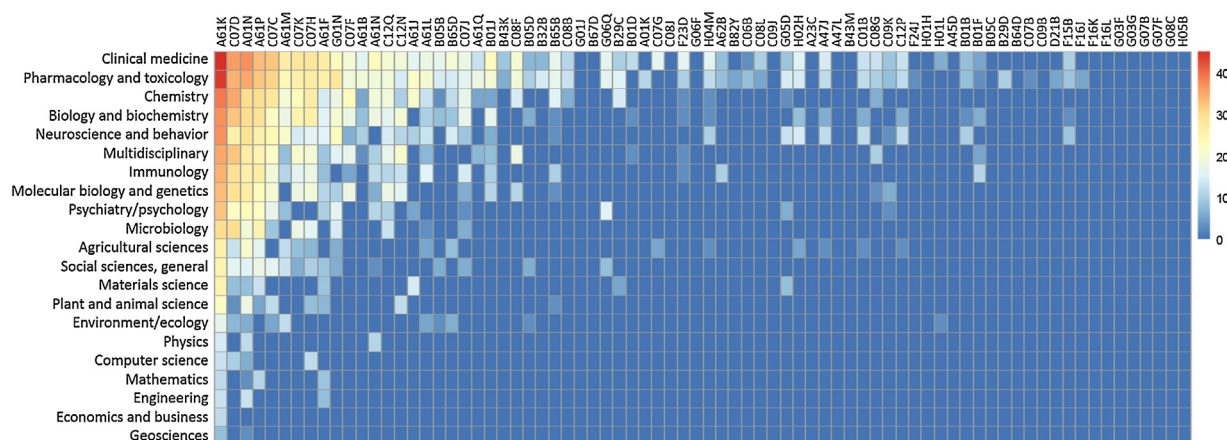
Fig. 9. Direct interaction of science and technology areas for drug innovation.

To map the interactions between scientific activities and technology inventions, we then match the ESI research areas of publications to 4-digit IPC technology classes based on scientific references in patents. This science-technology matrix is shown in Fig. 9, which is a heatmap in which science areas occupy one dimension and technology fields the other. The cells in such a science-technology matrix represent the intensity of the particular relation. The color of a cell indicates the intensity of the relation. The color scale runs from dark blue (low-intensity) via white and yellow to red (high-intensity). A high intensity reflects that a science field is frequently cited in a particular technology domain. The intensity linking science field i and technology field j is calculated as "the number of scientific citations from patents in technology j to publications in science domain i". As such a matrix contains information on the interaction of all science areas with all technology fields some of the combinations are unlikely to have a value. Publication from the science area 'clinical medicine' has the broadest influence on several technical fields. The most active science-technology interaction for drug innovation can be found in the research area of clinical medicine, pharmacology, chemistry/biochemistry and medical preparations technique (A61 K) and heterocyclic compounds technique (C07D). Fig. 6 also shows an orange cell that represents the link between the science field 'immunology'/'molecular biology/genetics' with the technical field 'medical preparations'. A yellow cell that represents the link between 'material science' and 'medical preparations' can also be detected. Our observation is in accordance with a recent study by (Malva, Kelchtermans, Leten, & Veugelers, 2015). They found immunology is a science field that is frequently cited in patents. This field receives most citations from three IPC classes: A61 (medical and veterinary technologies), C07 (organic chemistry), C12 (biochemistry).

3.8. The convergence of science and technology

With the ever increasing complexity of innovation, it is important to effectively combine knowledge from different fields or sources such as science and technology to create innovation which contains not only the integrated value but also synergies of the combined knowledge. The characteristics of scientific knowledge are different compared to those of technological knowledge. Both knowledge sources have distinguished characteristics and play distinctive roles in the invention process (Brooks, 1994). The main purpose of science is creating new knowledge and solving fundamental problems while developing scientific laws and theories that describe and explain the causes and effects of nature's phenomena. Therefore, output from scientific research is rarely directly applicable when releasing new product in the market. Even in scientific research-intensive industries like the chemical or pharmaceutical industries, the scientific knowledge from basic research institutes is difficult to apply right away (Vianen, 1990). On the other hand, technological knowledge is better suited to satisfying technological trends and market needs than scientific knowledge. In sum, science acts as exploratory action in R&D while technology aims at an effective recombination of existing knowledge and its practical improvement. By converging these two distinguished knowledge sources, new paradigms can spread.

Due to complementary roles and effects of science and technology in the invention process, the convergence of science and technology produces the synergies that lead to the development of more impactful innovation than processes purely depending on either science or technology. Especially, during the invention process, inventors can be inspired and stimulated by the convergence between cross-sources of knowledge (Brooks, 1994). To calculate convergence of science and technology, we adopts a measurement, namely *Convergence Ratio*, which was suggested by (Trajtenberg, Henderson, & Jaffe, 1997) and applied recently by (Lee, Park, & Kang, 2018). This variable represents the ratio of the scientific knowledge relative to the entire knowledge, both scientific and technological, that was used in innovation as described in the drug patent. It is calculated by the number of scientific publications over the total references of the focal patent. Since more than 80% of the focal patents leading to the new drugs innovation have scientific non-patent references, we try to investigate the effects of convergence of science and technology on pharmaceutical innovation.
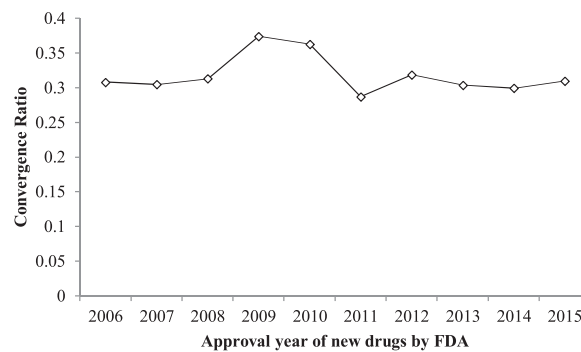
Fig. 10. Convergence Ratio of Science and technology for new drugs approved by FDA.
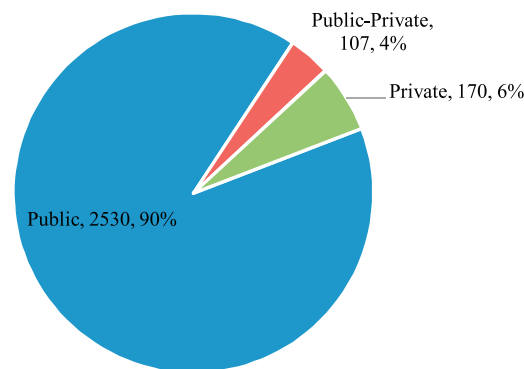
Fig. 11. Funding source of articles cited by drug patents.

We use focal drug patents' backward references as well as Web of Science/PubMed listed scientific publications in non-patent references to represent and measure convergence of science and technology. Specifically, we analyze the proportion of scientific knowledge in convergence for those successfully innovative drugs. In Fig. 10, a relative stable proportion (on average 0.33, 1/3) trend was observed for the convergence ratio of science and technology for new drugs approved by FDA. It seems that for leading to such type of successful innovation, the proportion of science and technology is 1:2. Science assists in solving difficulties in the invention process and, as a result, positively influences innovation. However, the objectives and aims of science mainly focused on solving fundamental issues, an overexploitation of scientific knowledge in the R&D process lead to solutions which are far from the demands of the technological market. This would lead to innovation which has less industrial impact than innovation derived from a balanced use of scientific (basic) and technological (applied) knowledge. In order to achieve impactful and successful innovation, it is important to understand the optimal combination and convergence of science and technology (Fig. 11).

3.9. Research funding of scientific publications cited by the drug patents

It is really a tough task to match all publications to their grant information, such as grant agency and granted research project number. When NLM first started to include grant numbers in MEDLINE citations in 1980, only agencies of the US Public Health Service (PHS) were taken. According to a recent investigation towards the number of PubMed records per year along with the percentage of those records containing various metadata, such as general funding acknowledgments and specific grant information, there are only 10%–15% articles containing specific grant information (K.W., B., R., & K., 2018). We captured all 14,831 PubMed articles in XML format, and found that 2567 articles (17.3%) were indexed with grant information (Table 3).

So, we turned to the 15,055 SCI-indexed articles and found that 2807 articles (18.6%) containing specific grant information. These 2807 granted articles were cited by 1227 patents (50.7%, 1227/2422). These 1227 patents were used to protect 295 drugs (48.8%, 295/605). So, the following analysis on grant sources can explain nearly half of patents and drugs concerned in this study. The origin of the 2807 granted articles which were afterwards cited by drug patents was assigned to two categories similar with the grouping framework for institutional source. The first category is the public sector, which comprises government funding agencies, universities, non-profit society and foundation. Private firms constitute the second category. We checked and cleaned the origin of the funding organizations for each article, and classified them into Public and Private. These papers acknowledged more than 130 different sources of funding support. The results show that a majority (90%) of papers are Public funded, suggesting that public funding plays an important role in the development of innovative

Table 3
Top 30 identified sources funding more than 1% articles cited by drug patents.

No.	Funding Organization	type	articles	%
1	NCI NIH HHS	public	577	20.6
2	NIDDK NIH HHS	public	308	11.0
3	NHLBI NIH HHS	public	298	10.6
4	NCRR NIH HHS	public	267	9.5
5	NIGMS NIH HHS	public	263	9.4
6	NIAID NIH HHS	public	233	8.3
7	NINDS NIH HHS	public	171	6.1
8	NIDA NIH HHS	public	165	5.9
9	NIA NIH HHS	public	125	4.5
10	NIMH NIH HHS	public	116	4.1
11	NEI NIH HHS	public	109	3.9
12	NICHD NIH HHS	public	104	3.7
13	NIADDK NIH HHS	public	95	3.4
14	PHS HHS	public	84	3.0
15	Merck & Company	private	83	3.0
16	NIAMS NIH HHS	public	71	2.5
17	NIH (only)	public	69	2.5
18	Sanofi	private	49	1.7
19	GlaxoSmithKline	private	48	1.7
20	Pfizer	private	47	1.7
21	Wellcome Trust	public	44	1.6
22	AstraZeneca	private	44	1.6
23	Johnson Johnson	private	44	1.6
24	Roche	private	42	1.5
25	Bristol Myers Squibb	private	39	1.4
26	NIEHS NIH HHS	public	37	1.3
27	NIDCR NIH HHS	public	36	1.3
28	Eli Lilly	private	34	1.2
29	Novartis	private	31	1.1
30	Abbott	private	27	1.0

drug products. Private-funded and Public-Private co-funded contribute to a smaller portion with 6% and 4%, respectively. As described in Table 1, the top two public and private funding sources identified were National Cancer Institute and Merck & Company, Inc., respectively.

In the first category, the majority of papers are funded by National Institute of Health and its subordinate bodies such as National Cancer Institute, National Institute of Diabetes and Digestive and Kidney Disease, National Heart Lung and Blood Institute, National Center for Research Resources, and so on. NIH funded 2377 papers, accounting for 85% of all funded papers. Other major public funding agency includes Public Health Service (PHS), Wellcome Trust (UK), Medical Research Council (UK) and Food and Drug Administration (USFDA). In the second category of industry funding, Merk & Company, Sanofi, GlaxoSmithKline, Pfizer, AstraZeneca, Johnson & Johnson, are the major funders.

4. Discussions and implications

The focus of our analysis lies on drugs approved for the United States, the world's largest market for pharmaceutical products. In this study, taking the objective data about the new drugs as the origin, a backward tracking model is used to measure the knowledge flow across drug products, their key patents, the cited scientific papers, and the research funding through citation analysis and linkage analysis. Several interesting observations and implications are suggested by in retrospect examining the scientific and technical knowledge which have been translated into successful innovation.

4.1. Developing regulatory science since the translational delay inherently exists

When talking about the time lag from research discovery to clinical practice, the most cited opinion is 16–17 years. Grant and Ioannidis JPA considered that it would be 16 to 17 years from basic research to new drugs approval (Contopoulos-Ioannidis, Alexiou, Gouvias, & Ioannidis, 2008; Grant, Green, & Mason, 2003). According to (Reidenberg & Erle, 2012)'s findings, the delay between basic research and applied results, measured as the median interval from the most recent basic paper to drug approval in the present study is at least 12 years for drug development after a drug target has been fully identified and validated. This is the inherent nature of new drug development and the natural laws of scientific research. Due to the innate translational delay between basic research and new drugs, it is necessary to improve the regulation science for shortening the drug development process. This study indicates that it will take about 10 years from the application of earliest patent to the approval of the new drug. Also such high-value patents in FDA Orange Book tend to cite scientific knowledge published on average 10–15 years ago. It should be recognized that the process of research and development of

new drugs is a long-term and uncertain process, and it is unrealistic to obtain huge returns in the short term. It is necessary to develop regulatory science since the translational delay inherently exists.

4.2. The important role of public science on pharmaceutical innovation

Our research builds on and extends previous work about the role of public science on innovation in several ways. Although a considerable body of research has identified many new drugs and vaccines approved by FDA were discovered through research carried out in public-sector research institutions (Stevens et al., 2011), and some authors have done so in the case of biotechnology (McMillan et al., 2000), and a bibliometric data analysis has shown scientific knowledge stemming from the public sector is found to be important for all different types of innovations (from radical to incremental) in the pharmaceutical industry (Sternitzke, 2010), ours adds to a long line not only by linking the new drugs to their basic patents, and drug patents to scientific articles but also by linking the articles back to funding sources.

From the perspective of knowledge flow, the present study showed that on average a new drug needs four or five basic patents to protect its active ingredient, formulation, or methods of use for approved indications, while one basic patent needs to cite on average six papers in the non-patent references. Therefore, an average of 24 scientific research papers is needed to support the new drugs innovation. The technology linkage of new drugs was relatively stable while the science linkage of technology inventions increased rapidly. Among the scientific papers cited by drug patents, private-institution originated papers are only a quarter of the public. By linking theses scientific papers with funding sources, we found a large majority (90%) are public-funded and only a very small part are private-funded or public-private joint-funded. Our study also indicates the importance of research on such fields as pharmacology, chemistry (including medicinal chemistry, biochemistry, and organic chemistry), molecular biology, neurosciences, and immunology on new drugs innovation. These findings indicate that the pharmaceutical innovation is even more public science linked than biotechnology industry compared with the data in (McMillan et al., 2000). That is, the distance between the science base and the innovation system is much closer in pharmaceutical fields than it is in other technological fields. This resonates with previous results (Sternitzke, 2010), which investigated the pharmaceutical industry which is, much more than other industries, dependent on scientific advances as well as work done in the public sector. The link between basic science and drug discovery has even increased over time, with companies moving from a mere random screening of compounds towards informed discovery methods fuelled by scientific knowledge (Malva et al., 2015)

Further, recombination of the scientific knowledge may produce radical inventions. In contrast to the conventional wisdom that radical inventions are based less on existing knowledge, (Schoenmakers & Duysters, 2010) find that radical inventions are induced by the recombination over more knowledge domains. The combination of knowledge from domains that might usually not be connected seems to deliver more radical inventions. Another research uncovered that radical innovations here build on more basic scientific knowledge than do incremental innovations, confirming that new technological trajectories are rooted in scientific discoveries (Sternitzke, 2010).

4.3. Basic research is similar with applied research for linking pharmaceutical innovation

There is no obvious relationship between "basicness" and linkages to the resulting patents' impact and to drugs innovation. Basic and applied research is quite similar in their contribution share of publications cited by patents, and the potential of resulting high-impact drug patents. One interpretation of this is that clinical research plays equally important role with basic research such as the fields of pharmacology and chemistry. The distribution of scientific papers by Essential Science Indicators Research Areas showed one third of scientific knowledge is from clinical medicine, the other one third is from pharmacology and chemistry, and the last one third is from biotechnology, immunology and other research fields. During drug development process in the pharmaceutical industry, additional patent filings as well as publishing activities of firms occur at about the same time, so firms foster their developmental activities once the basic patents are filed and marketing of the drug is about to start. The stream of papers following the basic patents is primarily clinical (Sternitzke, 2010). Our results argue that the basic/applied distinctions may not be so useful in thinking about what types of scientific research is more productive in resulting drug innovation.

The distribution of investments across the types of medical scientific research has been changed according to a data related survey by (Rd et al., 2015). Our further investigation uncovered that pharmaceutical companies shifted funding to late-phase clinical trials and away from discovery activity such as target identification and validation in pre-human/preclinical activities (only accounting for 17.2% in 2016). This shift toward clinical research and development reflects the increasing costs, complexity, and length of clinical trials but may also reflect a deemphasis of early discovery efforts by the US pharmaceutical industry. While industry has shifted funding to clinical trials, the share of National Institute of Health (NIH) contributions dedicated to basic research and applied research tend to be evenly divided, with 51% focused on basic research and 49% on applied research in 2017[4]. NIH defines applied research as research that directly involves human subjects or human tissues. It can be patient oriented research, epidemiological and behavioral studies, or outcomes research and health services research. From the lessons of the drug development of artemisinin, discovered by Youyou Tu who wins the Nobel Prize, we

---

[4] https://www.nsf.gov/statistics/fedfunds/.

understood that the successful pharmaceutical innovation is closely related to the capability of clinical research (Tang & Du, 2016). So, a balanced basic research and applied research maybe essential for fostering innovation drug innovation because it is a complete chain translating from basic discovery to clinical evidence then to clinical practice.

4.4. Establishing the public–private partnerships for pharmaceutical innovation

Historically, public-sector researchers have performed the upstream, basic research that elucidated the underlying mechanisms of disease and identified promising points of intervention, whereas corporate researchers have performed the downstream, applied research resulting in the discovery of drugs for the treatment of diseases and have carried out development activities to bring them to market. However, the boundaries between the roles of the public and private sectors have shifted substantially since the dawn of the biotechnology era, and the public sector now has a much more direct role in the applied-research phase of drug discovery, as is shown in our research. Results in this study also strated the characteristics of scientific papers cited by patents, (1) higher citation impact than the average level of all papers in the same field; (2) a large proportion collaborated with the industry. Our results are consistent with arguments that having an author affiliated with the pharmaceutical or biotechnology industry was associated with an eight- to 10-fold accelerated translational process. For 101 basic research papers published in the top 6 basic research journals from 1979 to 1983 which were claimed by their authors to be innovative in clinical treatment or healthcare prevention, the rate at which the first randomized (or first positive) trial was published was considerably faster when there was industry involvement (Contopoulos-Ioannidis, Ntzani, & Ioannidis, 2003). By linking patentable inventions and prior research, (Ahmadpoor & Jones, 2017) found that universities tended to cite their own research directly in their patents, but companies tend to rely on outsiders for their foundational research. Radical innovations and technological breakthroughs build on prior own scientific work to a higher degree than do market breakthroughs or incremental innovations, but this does not hold true for own prior technical work, i.e. own patents serving as the basis for new drugs. So absorbing externally generated knowledge works quite well within pharmaceutical firms, and conducting own scientific research is instrumental in this industry for generating radical innovations and technological breakthroughs (Sternitzke, 2010). The pharmaceutical innovation is characterized with multidisciplinarity, large capital expenditure, and a long cycle of R&D. The declining productivity of drug research and development is of major concern for private and public stakeholders in the pharmaceutical industry, and in health care more broadly (Paul et al., 2010). One strategy to tackle this challenge that has gained momentum in recent years is the establishment of precompetitive public–private partnerships to focus on issues that are too large for single organizations to effectively address alone, such as the development of biomarkers of drug toxicity (Denee et al., 2012).

Pharmaceutical companies generally cooperate with basic research institutions, but they do not focus on results, but provide long-term research funding, while seeing whether high-level scientists turn to work in enterprise to obtain long-term mutually beneficial cooperation. The R&D center of United States pharmaceutical companies gathers in Boston because Boston has the best universities, including Harvard and MIT, with the effect of high-end talent gathering. As a result, pharmaceutical companies collaborating with universities or research institutes and funding basic research to these public sectors will help to reduce risk and improve the chances of success. Also, organizations need to enhance their scientific capacity by employing more scientists who are familiar with scientific language as well as encouraging R&D towards more fundamental and basic principal to archive more impactful innovation. This calls for an investment in basic research and an increase in R&D collaborations with scientific institutions such as universities.

4.5. A balanced convergence of scientific and technological knowledge for drug innovation

We categorized an innovation's background knowledge into scientific and technological knowledge and empirically analyzed the converging scientific and technological knowledge for the resulting drug innovation. In such a knowledge intensive industry, new drug innovation is a complex process that combines knowledge with new ideas in a creative way from the knowledge side. In a recent study, (Lee et al., 2018a) use focal patents' backward references as well as SCI listed scientific publications in non-patent references to represent and measure convergence of science and technology. The results show that convergence of science and technology has a significant impact on innovation, and the effect varies with the ratio of scientific to technological knowledge. While the addition of scientific knowledge increases innovation impact when innovation is mostly based on technological knowledge, increasing the ratio of scientific knowledge beyond a certain point diminishes the influences of innovation impact, yielding a curvilinear relationship. In this study, we found that the so-called certain point maybe 1/3 for the successful pharmaceutical innovation as described by the new drugs approved by FDA. The results are consistent with the effects in previous studies which show that scientific searching activity has a positive effect on innovation (Grupp, 1996), which reconsidered the spillover effects and the science base of innovations. Therefore, it is important to maintain a balance between science and technology rather than overly reliance on one side.

For managers of pharmaceutical innovation organizations, the results of this study also present a suitable research strategy for their R&D activities. It provides inputs for a successful knowledge search strategy. In order to achieve successful impactful innovation, rather than focusing on only technology, convergence with science at moderate levels is important and that organizations should spread their search to cover both fundamental and basic fields as well as technological domains.

5. Limitations and future research

Although our analysis is a large-scale evaluation of a chain-linkage between funding, science, technology and innovation, there are several limitations. First, measuring direct linkages through patents citing articles may be also imperfect. Scientific references in patent documents can be used as indicators reflecting science-technology interactions. Whether they reflect a direct 'knowledge flow' from science to technology is subject of debate (Callaert, Pellens, & Van Looy, 2014). One important reason lies at the subject of patent citation behavior varying from inventors/applicants to examiners. (Li, Chambers, Ding, Zhang, & Meng, 2014) conclude that the non-self-citation by inventor/applicant is quite noisy and cannot indicate science linkage and that self-citation by inventor/applicant, although limited, is more appropriate for understanding science linkage. In addition, our measures of "basicness" only capture, imperfectly, some of the relevant dimensions in the age-old debates regarding basic versus applied research. Despite these limitations, we provide several new stylized facts on the translation and convergence by funding-science-technology-innovation linkages analysis. In further research, we will distinguish the non-patents references cited by inventor/applicant with examiner and also non-self-citation with self-citation, to map the new science-technology linkages and compared the results with this study.

Author contributions

Jian Du: Conceived and designed the analysis, Collected the data, Performed the analysis, Wrote the paper.
Peixin Li: Contributed data or analysis tools, Performed the analysis.
Qianying Guo: Contributed data or analysis tools, Performed the analysis.
Xiaoli Tang: None.

Acknowledgements

References

Ahmadpoor, M., & Jones, B. F. (2017). The dual frontier: Patented inventions and prior scientific advance. *Science*, *357*(6351), 583–587. http://dx.doi.org/10.1126/science.aam9527

Bowen, A., & Casadevall, A. (2015). Increasing disparities between resource inputs and outcomes, as measured by certain health deliverables, in biomedical research. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(36), 11335–11340. http://dx.doi.org/10.1073/pnas.1504955112

Boyack, K. W., Patek, M., Ungar, L. H., Yoon, P., & Klavans, R. (2014). Classification of individual articles from all of science by research level. *Journal of Informetrics*, *8*(1), 1–12.

Brooks, H. (1994). The relationship between science and technology. *Research Policy*, *23*(5), 477–486.

Callaert, J., Pellens, M., & Van Looy, B. (2014). Sources of inspiration? Making sense of scientific references in patents. *Scientometrics*, *98*(3), 1617–1629. http://dx.doi.org/10.1007/s11192-013-1073-x

Cambrosio, A., Keating, P., Mercier, S., Lewison, G., & Mogoutov, A. (2006). Mapping the emergence and development of translational cancer research. *European Journal of Cancer*, *42*(18), 3140–3148. http://dx.doi.org/10.1016/j.ejca.2006.07.020

Capaldo, A., Lavie, D., & Messeni Petruzzelli, A. (2017). knowledge maturity and the scientific value of innovations: The roles of knowledge distance and adoption. *Journal of Management*, *43*(2), 503–533. http://dx.doi.org/10.1177/0149206314535442

Carpenter, M. P., Cooper, M., & Narin, F. (1980). Linkage between basic research literature and patents. *Research Management*, *23*(2), 30–35. http://dx.doi.org/10.1080/00345334.1980.11756595

Cleary, E. G., Beierlein, J. M., Khanuja, N. S., McNamee, L. M., & Ledley, F. D. (2018). Contribution of NIH funding to new drug approvals 2010–2016. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(10), 2329–2334. http://dx.doi.org/10.1073/pnas.1715368115

Contopoulos-Ioannidis, D. G., Alexiou, G. A., Gouvias, T. C., & Ioannidis, J. P. A. (2008). Medicine: Life cycle of translational research for medical interventions. *Science*, *321*(5894), 1298–1299. http://dx.doi.org/10.1126/science.1160622

Contopoulos-Ioannidis, D. G., Ntzani, E. E., & Ioannidis, J. P. A. (2003). Translation of highly promising basic science research into clinical applications. *American Journal of Medicine*, *114*(6), 477–484. http://dx.doi.org/10.1016/S0002-9343(03)00013-5

Denee, T. R., Sneekes, A., Stolk, P., Juliens, A., Raaijmakers, J. A. M., Goldman, M., . . . & Janssen, J. W. (2012). Measuring the value of public–private partnerships in the pharmaceutical sciences. *Nature Reviews Drug Discovery*, *11*, 419. http://dx.doi.org/10.1038/nrd3078-c1 https://www.nature.com/articles/nrd3078-c1#supplementary-information

Duda, G. N., Grainger, D. W., Frisk, M. L., Bruckner-Tuderman, L., Carr, A., Dirnagl, U., . . . & Volk, H. D. (2014). Changing the mindset in life sciences toward translation: A consensus. *Science Translational Medicine*, *6*(264) http://dx.doi.org/10.1126/scitranslmed.aaa0599

Friedman, Y. (2010). Location of pharmaceutical innovation: 2000–2009. *Nature Reviews Drug Discovery*, *9*(11), 835–836. http://dx.doi.org/10.1038/nrd3298

Friedman, Y. (2017). Where are drugs invented, and why does it matter? *ACS Medicinal Chemistry Letters*, *8*(6), 589–591. http://dx.doi.org/10.1021/acsmedchemlett.7b00185

Glänzel, W., & Meyer, M. (2003). Patents cited in the scientific literature: An exploratory study of' reverse' citation relations. *Scientometrics*, *58*(2), 415–428. http://dx.doi.org/10.1023/A:1026248929668

Grant, J., Cottrell, R., Fawcett, G., & Cluzeau, F. (2000). Evaluating' payback' on biomedical research from papers cited in clinical guidelines: Applied bibliometric study. *British Medical Journal*, *320*(7242), 1107–1111.

Grant, J., Green, L., & Mason, B. (2003). Basic research and health: A reassessment of the scientific basis for the support of biomedical science. *Research Evaluation*, *12*(3), 217–224. http://dx.doi.org/10.3152/147154403781776618

Grupp, H. (1996). Spillover effects and the science base of innovations reconsidered: An empirical approach. *Journal of Evolutionary Economics*, *6*(2), 175–197.

Hu, Y., Scherngell, T., Man, S. N., & Wang, Y. (2013). Is the United States still dominant in the global pharmaceutical innovation network? *PLoS One*, *8*(11) http://dx.doi.org/10.1371/journal.pone.0077247

Huang, M. C., Fang, S. C., & Chang, S. C. (2011). Tracking R&D behavior: Bibliometric analysis of drug patents in the Orange Book. *Scientometrics*, *88*(3), 805–818. http://dx.doi.org/10.1007/s11192-011-0400-3

Jefferson, R. (2017). Comment: Turning science into social outcomes. *Nature*, *548*(7666), S8. http://dx.doi.org/10.1038/548S8a

Jefferson, O. A., Jaffe, A., Ashton, D., Warren, B., Koellhofer, D., Dulleck, U., . . . & Jefferson, R. A. (2018). Mapping the global influence of published research on industry and innovation. *Nature Biotechnology*, *36*(1), 31–39. http://dx.doi.org/10.1038/nbt.4049

Jibu, M. (2014). Mapping of scientific patenting: Toward the development of' J-GLOBAL foresight'. *Technology Analysis and Strategic Management*, *26*(4), 485–498. http://dx.doi.org/10.1080/09537325.2013.877129

Jones, D. S., Cambrosio, A., & Mogoutov, A. (2011). Detection and characterization of translational research in cancer and cardiovascular medicine. *Journal of Translational Medicine*, *9* http://dx.doi.org/10.1186/1479-5876-9-57

K.W, B, & R., K. (2018). Accurately identifying topics using text: Mapping PubMed. In *STI 2018 Conference Proceedings*. pp. 107–115.

Lewison, G., & Paraje, G. (2004). The classification of biomedical journals by research level. *Scientometrics*, *60*(2), 145–157. http://dx.doi.org/10.1023/B:SCIE.0000027677.79173.b8

Lee, C., Park, G., & Kang, J. (2018a). The impact of convergence between science and technology on innovation. *Journal of Technology Transfer*, *43*(2), 522–544. http://dx.doi.org/10.1007/s10961-016-9480-9

Lee, C., Park, G., & Kang, J. (2018b). The impact of convergence between science and technology on innovation. *Journal of Technology Transfer, 43*(1), 1–23.

Li, D., Azoulay, P., & Sampat, B. (2017). The applied value of public investments in biomedical research. *Science*, *356*(6333), 78–81. http://dx.doi.org/10.1126/science.aal0010

Li, R., Chambers, T., Ding, Y., Zhang, G., & Meng, L. (2014). Patent citation analysis: Calculating science linkage based on citing motivation. *Journal of the Association for Information Science & Technology, 65*(5), 1007–1017.

Liu, W., Hu, G., & Tang, L. (2018). Missing author address information in web of science – An explorative study. *Journal of Informetrics*, *12*(3), 985–997. http://dx.doi.org/10.1016/j.joi.2018.07.008

Malva, A. D., Kelchtermans, S., Leten, B., & Veugelers, R. (2015). Basic science as a prescription for breakthrough inventions in the pharmaceutical industry. *Journal of Technology Transfer*, *40*(4), 670–695. http://dx.doi.org/10.1007/s10961-014-9362-y

McMillan, G. S., Narin, F., & Deeds, D. L. (2000). An analysis of the critical role of public science in innovation: The case of biotechnology. *Research Policy*, *29*(1), 1–8. http://dx.doi.org/10.1016/S0048-7333(99)00030-X

Narin, F., & Noma, E. (1985). Is technology becoming science? *Scientometrics*, *7*(3–6), 369–381. http://dx.doi.org/10.1007/BF02017155

Narin, F., Pinski, G., & Gee, H. H. (1976). Structure of the biomedical literature. *Journal of the American Society for Information Science*, *27*(1), 25–45. http://dx.doi.org/10.1002/asi.4630270104

National Institute of Health, O. o. P. A. (2014). *IDENTIFYING TRANSLATION: iTrans* Retrieved from. https://www.dpcpsi.nih.gov/sites/default/files/iTrans_one_pager%2004172018.pdf

Paul, S. M., Mytelka, D. S., Dunwiddie, C. T., Persinger, C. C., Munos, B. H., Lindborg, S. R., . . . & Schacht, A. L. (2010). How to improve R&D productivity: The pharmaceutical industry's grand challenge. *Nature Reviews Drug Discovery*, *9*, 203. http://dx.doi.org/10.1038/nrd3078 https://www.nature.com/articles/nrd3078#supplementary-information

Rd, M. H., Matheson, D. H., Cairnssmith, S., George, B. P., Palisch, C., & Dorsey, E. R. (2015). The anatomy of medical research: US and international comparisons. *Journal of the American Medical Association, 313*(2), 174–189.

Reidenberg, M. M., & Erle, H. (2012). U.S. Basic research: Delayed drug development. *Science*, *337*(6102), 1605. http://dx.doi.org/10.1126/science.337.6102.1605-a

Roach, M., & Cohen, W. M. (2013). Lens or prism? Patent citations as a measure of knowledge flows from public research. *Management Science*, *59*(2), 504–525. http://dx.doi.org/10.1287/mnsc.1120.1644

Schoenmakers, W., & Duysters, G. (2010). The technological origins of radical inventions. *Research Policy, 39*(8), 1051–1059.

Sternitzke, C. (2010). Knowledge sources, patent protection, and commercialization of pharmaceutical innovations. *Research Policy*, *39*(6), 810–821. http://dx.doi.org/10.1016/j.respol.2010.03.001

Stevens, A. J., Jensen, J. J., Wyller, K., Kilgore, P. C., Chatterjee, S., & Rohrbaugh, M. L. (2011). The role of public-sector research in the discovery of drugs and vaccines. *New England Journal of Medicine, 364*(6), 535–541.

Sung, H. Y., Wang, C. C., Huang, M. H., & Chen, D. Z. (2015). Measuring science-based science linkage and non-science-based linkage of patents through non-patent references. *Journal of Informetrics*, *9*(3), 488–498. http://dx.doi.org/10.1016/j.joi.2015.04.004

Tang, X., & Du, J. (2016). The performance of China's biomedical innovation: A scientometric analysis. *Science China Life Sciences*, *59*(10), 1074–1082. http://dx.doi.org/10.1007/s11427-016-5078-6

Tang, L., Hu, G., & Liu, W. (2017). Funding acknowledgment analysis: Queries and caveats. *Journal of the Association for Information Science and Technology*, *68*(3), 790–794. http://dx.doi.org/10.1002/asi.23713

Thonon, F., Saghatchian, M., Boulkedid, R., & Alberti, C. (2015). Using bibliometrics to measure the impact of cancer research on health service and patient care: Selecting and testing four indicators. In *Paper Presented at the Proceedings of ISSI 2015 Istanbul: 15th International Society of Scientometrics and Informetrics Conference.*

Trajtenberg, M., Henderson, R., & Jaffe, A. (1997). University versus corporate patents: A window on the basicness of invention. *Economics of Innovation & New Technology, 5*(1), 19–50.

Vale, R. D., & Hyman, A. A. (2016). Priority of discovery in the life sciences. *eLife*, *5*, e16931 http://dx.doi.org/10.7554/eLife.16931

Vianen, B. G. V. (1990). An exploration of the science base of recent technology. *Research Policy, 19*(1), 61–81.

Wang, J., & Shapira, P. (2011). Funding acknowledgement analysis: An enhanced tool to investigate research sponsorship impacts: The case of nanotechnology. *Scientometrics*, *87*(3), 563–586. http://dx.doi.org/10.1007/s11192-011-0362-5

Weber, G. M. (2013). Identifying translational science within the triangle of biomedicine. *Journal of Translational Medicine*, *11*(1) http://dx.doi.org/10.1186/1479-5876-11-126