

# Exploring temporal relationships between scientific and technical fronts: a case of biotechnology field

Mu-Hsuan Huang · Ssu-Han Chen · Chia-Ying Lin · Dar-Zen Chen

Received: 29 March 2013  
© Akadémiai Kiadó, Budapest, Hungary 2013

**Abstract** Biotechnology is an expanding interdisciplinary field in which the interactions of science and technology (S&T) are more and more intensified. Question raised regarding the dynamic interactions between S&T encourages us to propose a series of methodologies for examination. Using high-impact publications and patents as the proxy measures, two document sets are transformed into the scientific and technical front trajectories respectively, and then each subject is categorized into either basic science, or applied technology, or co-existence. The results show that, in the biotechnology field, subjects of embryonic or mesenchymal stem cells, RNA interference, microRNA, and microbial fuel cell are in the basic science phase; those of plant breeding, seed diversity, and taste receptors have been applied to practice. There also exists interactions between S&T in the subjects of disease treatment and gene analysis platform, in which the behavior of technology precedes science, science precedes technology, or synchronous development can be observed.

**Keywords** Bibliometric · Patentometric · Scientific front · Technical front · Trajectory · Biotechnology

## Introduction

Since industrial revolution, unidirectional viewpoint of the relationships between science and technology (S&T) has been replaced by collaboration. Scientific research is conducted

---

M.-H. Huang  
Department of Library and Information Science, National Taiwan University, Taipei, Taiwan, ROC

S.-H. Chen · D.-Z. Chen (✉)  
Department of Mechanical Engineering, Institute of Industrial Engineering, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, 10617 Taipei, Taiwan, ROC  
e-mail: dzchen@ntu.edu.tw

C.-Y. Lin  
Department of Mechanical Engineering, National Taiwan University, Taipei, Taiwan, ROC

for application based on modern instruments for analysis, and is used extensively to better technologies (Gardner 1999; Bassecoulard and Zitt 2005). While scientists induce systematic observations and then deduce new knowledge or construct fundamental theories, technologists design, create, or improve materials, functions of product, and manufacturing procedure (Gardner 1999). The knowledge-driven society has made the path to knowledge creation more complicated than ever.

The biotechnology is a rapidly expanding interdisciplinary field. Recently, it has been noted that S&T are actively interconnected (Arber 2009; Shalhevet et al. 2001). The United Nations Convention on Biological Diversity defines biotechnology as “any technological application that uses biological systems, living organisms, or derivatives thereof, to make or modify products or processes for specific use.” Biotechnology applies scientific and engineering principles to the processing of materials through biological systems as well as living organisms, especially in the context of pressing global challenges in the field of agriculture, food, medicine, plant, genomics, immunology, pharmaceutical therapy, diagnostic test, or other specific purposes (Gavrilescu and Chisti 2005). The future of biotechnology is optimistic (Braun 2002). Governments worldwide invest in this field and encourage research and development (R&D) in private companies and research institutes. In recent years, the connection between S&T is intensified as innovation cycle time shortens (Narin et al. 1997), i.e. progresses of S&T are no longer separately confined. Companies with application researchers on board indicate potential technological development. Scientists attempt to clarify the boundaries of specific bio-domains in terms of companies who would like to gain returns from them. Question is raised regarding the interaction between S&T in the field of biotechnology.

Publications and patents as information platforms have analogous features which are ideal “proxy measures” to represent the contents of S&T respectively (Verbeek et al. 2002). The quantitative appraisals of S&T relation made progress in the last two decades, which can now be employed to construct high-quality roadmaps to integrate the two respective fields (Kostoff and Schaller 2001). There are two main methods available: citation relation and lexical (content-based) relation. Citation relation exploits S&T relation, which is a citation from a patent (publication) to a non-patent (non-publication) reference Huang (2009). Counting the strength of linkage is then used to reveal that the impact of the knowledge resources have on technology development and vice versa. In general, most researchers of this area traced citations of publications in patents. Narin et al. (1997) revealed the increasing linkage phenomenon between US technology and public science. Narin and Olivastro (1998) investigated non-patent references on United States Patent & Trademark Office (USPTO) and European Patent Office (EPO) patents, showing large dependence of industries on public science. Verbeek et al. (2002) developed a citation linkage scheme to discern between some fields of technology that are highly science-interactive. Ribeiro et al. (2010) showed that scientific content of technology is continuously increasing, but the increasing degrees differ across countries and developing levels. Lo (2010a, b) revealed the knowledge resources from research had significant impact on technology development in the field of genetic engineering. Callaert et al. (2012) delineated the scientific footprint in technology in databases of USPTO and EPO. Unlike most contributions, Glänzel and Meyer (2003) did not use non-patent reference but explore citations of patents in publications. They found that the publications in chemistry-related fields were more likely to cite patents more than others. The counterpart, lexical relation, shows the lexical properties and connections between two realms from volumes of corpora by analyzing term frequency and (groups of) document similarity. Lexical relation has gradually been employed by researchers for its accessibility and thus be adopted in this

study. Co-word maps that mix publications and patents were presented by Engelsman and van Raan (1994) to characterize the S&T relation. Bassecoulard and Zitt (2005) explored the ability to establish correspondence tables between patent classification and scientific categories based on the chemical abstracts database. Lissoni and Montobbio (2008) used text-mining techniques to identify patent-publication pairs of a sampling of Italian scientists. Shibata et al. (2011) discovered the research subfields where there are scientific activities but no technological applications.

The S&T relationship described above indicates current topics and inter-activities among publications and patents in a specific field (Lissoni and Montobbio 2008; Nakamura et al. 2010; Shibata et al. 2011). Such understanding provides a bidirectional viewpoint for authors to gain insights on the development of a field. However, such results provide merely snapshot track, hence fail to capture the dynamics of their reciprocal inferences. In this study, we refer to Gardner, who outlined four categories for the nature of the S&T relation: the demarcationist view (separate fields), the idealist view (science precedes technology), the materialist view (technology precedes science), and the interactionist view (two-way interaction) (Gardner 1999). Gardner's categorization can be integrated with investigation of the S&T relationship by employing the temporal information. Drawing into the dynamics allows analyzers to form scientific or technical front trajectories and to identify the differences or interactions in development between S&T in a specific field. Extensive literature review, however, shows that there has been far less studies on the connection between scientific and technical front trajectories. Therefore, an integrated solution is designed to provide insights into this subject of debate. The expended resolution of exploring S&T gap can provide insights on the differences in more details.

The rest of this paper is as follows: the methodology is described and explained in Section "Research methodology"; the experimental environment is depicted and the results are presented in Section "Experiment results"; the concluding remarks and further suggestions are discussed in Section "Conclusions and further suggestions".

## Research methodology

The number of publications and patents needed for analysis is extensive to the extent that we cannot rely upon manual reading. This paper aims to provide a quantitative study on S&T datasets, observing evolution of scientific and technical fronts and their temporal relationships. The concept of "front" can be generally defined as a cluster of high-impact documents in which the documents are linked by the similarity measurements (Small and Griffith 1974; Glänzel and Czerwon 1996; Boyack et al. 2011). In this study, a front represents a cluster of high-impact documents which are lexical-connected. If the document set is composed of high-impact publications, the clusters are called scientific fronts. On the other hand, if the document set turns into high-impact patents, the corresponding clusters are technical fronts. The procedure of front detection is preceded in a network level where nodes are represented by publications or patents and edges are represented by text similarities. A text similarity is evaluated by the Salton's cosine of pairwise term vectors (Salton and McGill 1983). The analysis uses a self-programming toolkit under the R environment with the *igraph*, *tm*, *RWeka*, *stringr*, *openNLP*, *wordnet*, *plotrix*, *Snowball*, *proxy*, and *gdata* packages (<http://cran.r-project.org/>).

Note that this study is one of the extensions of our previous paper, in which Chen et al. (2012a, b) showed the evolution history of a technology field both domestically and internationally. More importantly, the status of the country in the global context was

simultaneously evaluated. Such methodology can be further applied to the analysis of the gap between two countries, two institutions, two laboratories, or other two different specialties. The analyzed sources can be publications as well as patents in order to examine the S&T relationship, which is the primary focus in this study. The methodology presented in Chen et al. (2012a, b), however, poses several inherent challenges. First, a high percentage of high-impact documents would be isolated and become neglected in network analysis because citation-link matrices of bibliographic coupling analysis are extremely sparse, i.e. the document coverage rate is low (Glänzed and Thijs 2011). A second and potentially more important problem is the fact that a fully content-based procedure is more pure and fluent than that of hybrid of citation and text information in this study. With a series of careful text manipulations procedures, each document is transformed into term-document vector, and similarities between document pairs are then evaluated, which could reduce the possibility of missing high-impact documents. The lexical relations between two distinct corpora, publications and patents, could also be studied using similar term-document vectors.

#### Showing evolving fronts by timeline plot

Data were retrieved. Timeline plot of front evolution is categorized into four stages. First, the high-impact documents are selected in terms of citation counts. Length and step size of the dynamic sliding window is then determined such that the fulltime dataset can be split up into a series of overlapping buckets. Second, the text similarities among document pairs in each bucket are evaluated. Third, the text similarity matrix in each bucket is individually clustered to detect fronts. Finally, strings are linked for qualified front pairs between successive buckets. Front trajectories are laid out on timeline to show S&T evolution. The above procedure is basically similar with that of Chen et al. (2012a, b), except text information is used to measure the similarity between document pairs. The detailed process is explained as follow:

#### *Selection of documents and determination the sliding window*

High-impact documents that receive more forward citations usually represent higher applicability, higher quality and greater impact on offspring. This study applies the method of Wu et al. (2010) which select high-impact documents in order to avoid disciplinary bias of average citation frequencies. High-impact documents are defined as those documents that are cited more than average and a standard deviation of cited counts in the same year. The window is defined through splitting the analyzed period into equidistant slices (Falkowski 2009) and sliding in overlapped perspective. Therefore, the addition or deletion movements of document are illustrated to simulate the dynamic environment (Lee and Chien 2013).

#### *Determination of document pairs with strong text similarities*

Text-based document analysis is one of commonly used approaches to measure similarity of document pairs. Since the titles of the documents and corpus condense most of salient contributions, the initial stage is to collect documents' titles and abstracts as a corpus. These unstructured data are then transformed into structure data, i.e. term vectors, using text mining. Specifically, terms from each document are extracted through lower case

conversion, multi-space stripping, number and punctuation removal, text singularization, and stopword elimination, and finally,  $l$ -level term hierarchy generation. In the hierarchy, a term at a higher level is lexically more general than that at a lower level. The length of term is truncated at the  $l$ -level in which the frequency of term is scarcely greater than one in the  $(l + 1)$ -level or in a higher level (Chen et al. 2012a, b). Count-based method is then used to form the term-document matrix and the similarities between documents are evaluated using Salton's cosine. Since partial similarities are extremely weak, a relative threshold introduced by Chen et al. (2013) is used to preserve relatively strong similarities. The relative threshold is defined by the average and standard deviation of the Salton's cosine.

*Detection of fronts over time*

With the  $N$  nodes (publications or patents) and  $M$  edges (similarities), a network can be composed by a  $N \times N$  adjacency matrix where its element ( $s_{ij}$ ) represents the Salton's cosine between the document pair of  $i$  and  $j$  and  $i, j = 1, 2, \dots, N$ . In network analysis, the fronts are detected using the weighted Girvan–Newman (GN) algorithm (Newman 2004), as it does not involve human judgments in setting the number of fronts in advance and is suitable for detecting clustering structure in weighted networks. The heuristic of weighted GN algorithm is: Weighted betweenness of edges are calculated. Remove the edge that has the highest betweenness from the network and recalculate betweenness of edges on the pruned network. We then evaluate modularity (Q) to represent the goodness of division:

$$Q = \frac{1}{2M} \sum_{ij} \left( s_{ij} - \frac{n_i n_j}{2M} \right) \delta(p_i, p_j) \tag{1}$$

where  $\delta(p_i, p_j)$  is 1 if documents  $i$  and  $j$  are in the same front and 0 otherwise;  $n_i$  is the number of lexical similar documents of document  $i$ . This process is repeated until edges are all removed. After iterations, the clustered result is determined based on the pruned network with highest modularity.

After GN clustering is carried out in one snapshot, size, normalized size, average age, and key terms are evaluated for each front. Among those features, size is number of documents in a front which reflects the influence of scope. Normalized size is in a percentage representation through division of size by the total number of documents in the network. Normalized size enables us to determine a level, called dominant front threshold ( $\alpha \in [0, 1]$ ), to filter weak fronts with relatively fewer documents involved. The average age reflects the age of a front in a snapshot. Finally, key terms of fronts are extracted to interpret their thematic topic. In this study, terms of each front are weighted by term frequency-inverse document frequency ( $tf-idf$ ) to measure the frequency and uniqueness of terms in a front compared with the others. The terms associated with the top  $tf-idf$  values in each front are then regarded as key terms (Chen et al. 2012a, b).

*Linkage of front string and presentation of timeline plot*

We described the continuity of fronts from one snapshot to another when a window slices over data overlapping in successive time. A front string is formed between fronts of two successive snapshots if they share at least one common document (Small 2006). Scientific or technical front evolutions are then visualized in a two-dimensional timeline in which fronts are plotted as a function of size and average age and time.

Detecting temporal relationships between scientific and technical fronts

Another main goal of this study investigates the interaction of scientific and technical fronts and transforms the problem into comparing the temporal relationships between them. As shown in Fig. 1, we suggest three categories to represent the relations between the scientific fronts and the technical ones in terms of their existence and temporal gaps. Type A indicates a front exists in the timeline plot of science but does not in that of technology. The front’s focuses may be too difficult to commercialize in current environment. By contrast, type B refers to front that does not exist in the timeline plot of science but does exist in that of technology. This indicates that the front could be regarded as an overly applied science or reflects a possible lack of science. Type C defines a front that co-exists in the timeline plots of both S&T contexts. In this situation, we expect the underlying fronts to have both reciprocal influence in which the temporal gap measure depends on the difference in average age of a scientific front and a technical front. Temporal gaps larger than, equal to, and smaller than zero correspond to technology preceding science (Type C-1), development being synchronous (Type C-2), and science preceding technology (Type C-3) respectively.

To check whether a front exists in the timeline plots of S&T, the measure of cosine similarity (CS) was employed to measure the lexical S&T connection. If  $x_{ji}$  is the value of the binary variable for the word  $i$  in front  $j$ , the CS measurement of the similarity between a scientific front  $s$  and a technical front  $t$  could be computed as follows:

$$CS(s, t) = \sum_i x_{si}x_{ti} / \sqrt{x_{si}^2} \sqrt{x_{ti}^2} \tag{2}$$

Theoretical values of CS are in the continuous [0, 1] range. A front matching threshold ( $\beta \in [0, 1]$ ) is set to determine whether the match of a front pair is qualified or not.

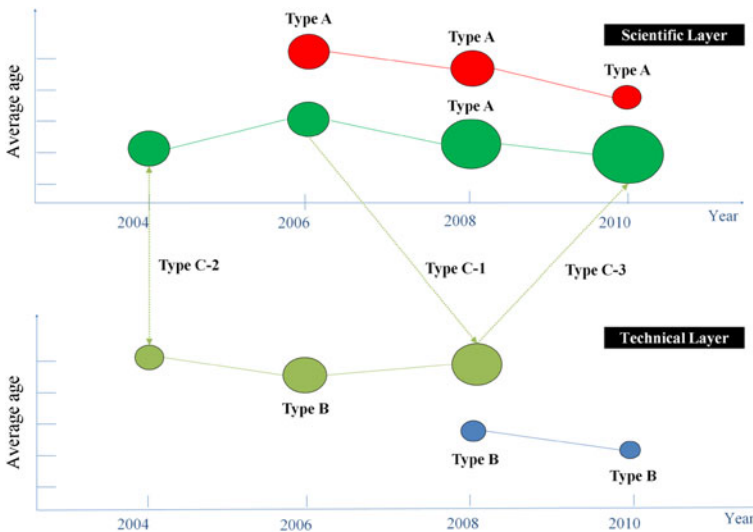


Fig. 1 Illustration of relations between science and technology

## Experiment results

### Case profile

Biotechnology has established itself as an emerging field in the twenty first century. It has long been utilized in knowledge management to assess competitive position and to avoid infringement. Publications and patents are used for analysis in this study to monitor the development of S&T and evaluate the position of peers in the market. The subject categories defined by the Institute for Scientific Information (ISI) are used to indicate different science fields. Biotechnology and applied microbiology is the subject category used in this paper. The United States Patent Classification (USPC) categories are used to indicate different technology fields. Patents were chosen in the NBER subcategories 33 (biotechnology as part of the drugs and medical category) which is represented by parts of the current main USPC 435 (chemistry: molecular biology and microbiology) and 800 (multicellular living organisms and unmodified parts thereof and related processes) (Hall et al. 2001). The time range spans from January 1, 2001, to December 31, 2010, with 203,478 publications and 54,080 issued utility patents retrieved respectively from the database of the Web of Science® (WOS) and the USPTO respectively after querying by the subject category and the USPC stated above.

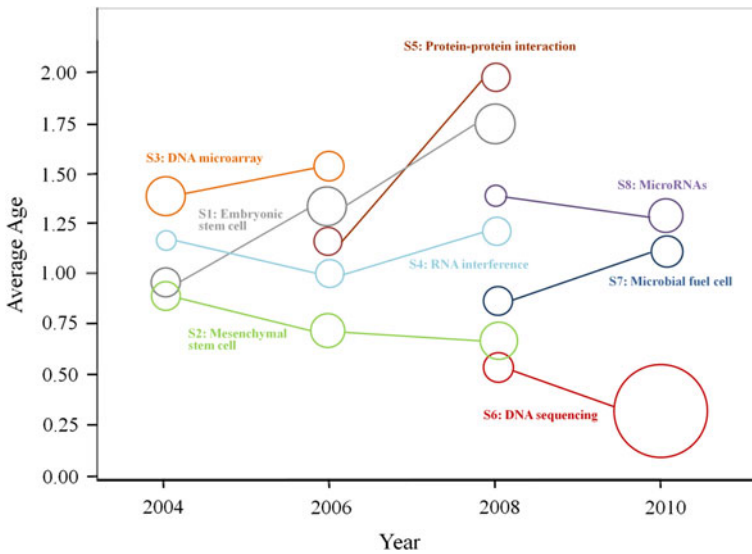
### Evolutions of publication and patent in biotechnology

High impact documents are selected using average cited counts and its standard deviation annually. There were a total 5,617 high-impact publications (2.76 %) and 2,902 high-impact patents (5.37 %). We divided the publications and patents into 4-year window to emphasize the dynamic of evolution and increase cluster accuracy. Falkowski's rule of thumb is employed to set a window overlap of 50 percent to decrease fluctuations of the rolling clustering (Falkowski 2009). Then high-impact documents and relatively strong similarities occurring in each snapshot aggregated as a publication or patent network which was then assembled as scientific or technical fronts through a GN algorithm. After the detection of fronts, dominant fronts were identified based on the Rosin thresholding (Perng and Chen 2011). The thresholds were set to 0.025 and 0.018 for the publication and patent sets respectively. Scientific and technical fronts in the biotechnology field covers eight evolving trajectories, marked as S1 to S8 and T1 to T8 as put forward in Figs. 2 and 3. Then those trajectories were post-assigned concise and descriptive names by human experts with the help of natural language processing.

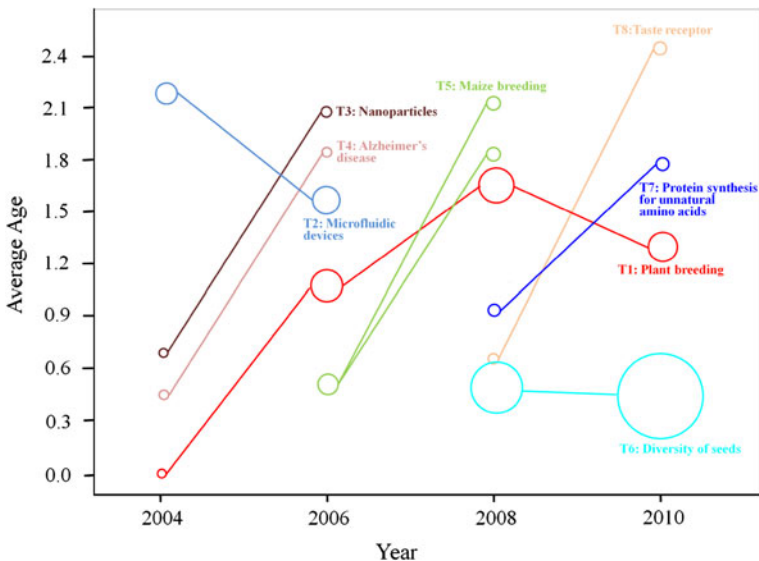
### *Results of the scientific front trajectories*

The eight scientific front trajectories are as follows:

- Embryonic stem cell (ESC) (S1): ESCs are derived from the inner cell mass of the blastocyst, an early-stage embryo. The characteristics of ESCs were verified in 2004. Pluripotency is one type of EDCs that are able to differentiate derivatives into three primary germ layers, which include more than 220 cell types. Additionally, under particular conditions, ESCs are capable of propagating themselves indefinitely. In 2006, molecular biology was used to discuss the change of participating proteins and genes during differentiation. The techniques and conditions selected for derivation and culture of ESC result in specific derivative cells. Because of the plasticity and



**Fig. 2** Presentation of scientific front trajectories in biotechnology



**Fig. 3** Presentation of technical front trajectories in biotechnology

potentially unlimited capacity for self-renewal, ESCs were proposed for use in regenerative medicine and tissue replacement after injury or disease, as well as other applications in 2008.

- Mesenchymal stem cells (MSC) (S2): Adult stem cells found in bone marrow, baby cord blood and blood in adult bodies, including MSC and hematopoietic stem cells. MSCs are studies to treat diseases such as cancer and regulation in gene therapy in



2004. Moreover, cell therapies, tissue engineering and regenerative medicine MSCs have been in development and clinical use in 2006. MSCs have self-renewal and proliferation capacities. Generally, a MSC can be divided into two cells, a mesenchymal stem cell and a cell with variety of types, including osteoblasts, adipocytes, and chondrocytes as well as myocytes and neurons. The study of these characteristics was developed through cellular and animal experiments in 2008.

- DNA microarrays (S3): DNA microarrays, known as “gene chips”, are able to contain tens of thousands of probes to measure the expression levels of large numbers of genes simultaneously or to identify multiple regions of a genome. Microarrays supporting bioinformatics are very complex and handle large volumes of information. In 2004, gene expression profiling, comparative genomic hybridization, SNP detection and alternative splicing detection became widely applied in microarray systems. Microarray-based gene expression profiling were used to identify genes whose expression changes in response to pathogens or other organisms by comparing gene expression in infected to that in uninfected cells or tissues in 2006.
- RNA interference (RNAi) (S4): RNAi is a process within living cells that moderates the activity of their genes. Interference technology was actively studied in 2004. The section of RNA or certain functions in living cells become inactive due to short interfering RNA (siRNA), cleaved from double-stranded RNA (dsRNA) out of the cell, binds to other specific messenger RNA (mRNA). In 2006, methods of transfection of siRNA into cells using viral carriers, antibodies or lipin gables was employed in order to treat disease. In 2008, compatibility of specific treatments improved, which made possible side-effects of the gene therapy a subject of interest.
- Protein–protein interaction (S5): This consists of two or more proteins binding together, often carrying their biological function. The protein interaction network system is built according to biochemistry, quantum chemistry and data statistics in order to analyze the relationship between a protein and another protein. This topic focused on the proteins in the human body in 2006. In 2008, information about diseases such as cancer, obtained from protein interaction network, was continuously researched, and may be applied to the design of medicines in the future.
- DNA sequencing (S6): This includes methods and technologies that are used to determine the order of the nucleotide bases—adenine, guanine, cytosine, and thymine—in a molecule of DNA. Through the development of algorithms and DNA-sequencing instruments, there has been an increase in the capability to read length of sequences per unit time. 454 pyrosequencing, illumina sequencing, SOLiD sequencing were applied for commercial purposes in 2008. During the development of next generation sequencing, types of cancer in DNA sequences were disclosed and compared to common cells in the body in 2010. In order to reduce the charge on DNA sequencing and to realize the differences between all genomes with cancer and enable the design of personalized drugs,
- Microbial fuel cells (MFC) (S7): MFCs are devices that can convert organic matters into electrical energy or fuel, and can also be used for disposal of wastewater. The development potentiality of MFC was discussed in the investigation of microorganisms around 2008. Generally, the reactions for microbial fuel cells are complex in anaerobic conditions. In 2010, a great deal of studies focused on the path of electron transfer, types of microorganism, and additional catalysis, raising the efficiency of microbial fuel cells.
- MicroRNAs (S8): These are short RNA molecules that bind to complementary sequences on target mRNA and result in target degradation or translational repression

and gene silencing. An available miR2Disease database that could interpret the relationships between miRNA and human disease was published in 2008, based on high-throughput sequencing and bioinformatics analysis. Recently, several miRNAs have been discovered to be linked with other types of cancer. This provides a way to prevent, diagnose, detect or treat miRNAs.

### *Results of the technical front trajectories*

The eight technical front trajectories are as follows:

- **Plant breeding (T1):** Plant breeding is the science of changing the genetics of plants in order to produce desired characteristics. It has been practiced for thousands of years to produce improved plants more easily. In order to maintain characteristics of variety, the maize hybrid variety was forced to change to inbred maize in agriculture by manual control in 2004. With the development of plant breeding, the characteristics of hybrid strains were discussed in 2006. Cotton cultivars are bred by the breeding of hybrid strains with the choice of high production and good properties in different varieties of cotton. According to transgenic technology, certain plants, such as canola and celery cultivars, possess the pest resistant genics and other desirable traits. The transmitting genics to the next generation through inbreeding or hybrid breeding is known as “genetic modification,” which was developed in 2008. The soybean cultivar bred by breeding of a hybrid strain and transgenic technology in 2010 improved breeding speed, and pest resistance compatibility.
- **Microfluidic devices (T2):** Microfluidic devices provide channels affected by surface tension and high specificity of chemical and physical properties for experiments and applications, where liquid or gas can flow in macro-scale system. They have the capability to purify cells, biological molecules, toxins and other materials. In 2004, the primary major of development of this technique was to improve microfluidic structure and integrate different models, such as pumps, valves, and channels, into a chip. With fast and precise control on the chip, microfluidic devices were furthered applied to enzymatic analysis, DNA analysis, DNA sequencing and proteomics medical examination and cell culture in 2006. Some papers have also studied the characteristics of chemical and physical properties for liquids or gases in microfluidic devices.
- **Nanoparticles (T3):** These provide a method to detect nucleic acids. The method involves the contact of the nucleic acid with one or more types of particles with oligonucleotides attached. In 2004, results of the hybridization of the oligonucleotides on the nanoparticles to the nucleic acid were obtained from a detectable change, which could be detected by silver staining. Moreover, specific DNA fragment assembly was been designed by nanofabrication utilizing nanoparticles, and was used in medicine in 2006.
- **Alzheimer’s disease (AD) (T4):** AD has been identified as a protein misfolding disease, caused by massive accumulation of the folded beta-amyloid (also written as A $\beta$ ) and tau proteins in the brain. Beta-amyloid, a fragment from a larger protein, is called an amyloid precursor protein (APP), a transmembrane protein that penetrates through the neuron’s membrane. In 2004, numerous ways to modify APP proteins, associate them with nucleic acids, and control beta-amyloid were to identify candidate therapeutics for the treatment or prevention of Alzheimer’s disease. In 2006, the important matter,  $\beta$ -secretase, was known as an enzyme to accelerate the division of APPs. The drugs to

block them based on the theory that inhibitive activity would prevent the building of beta-amyloid.

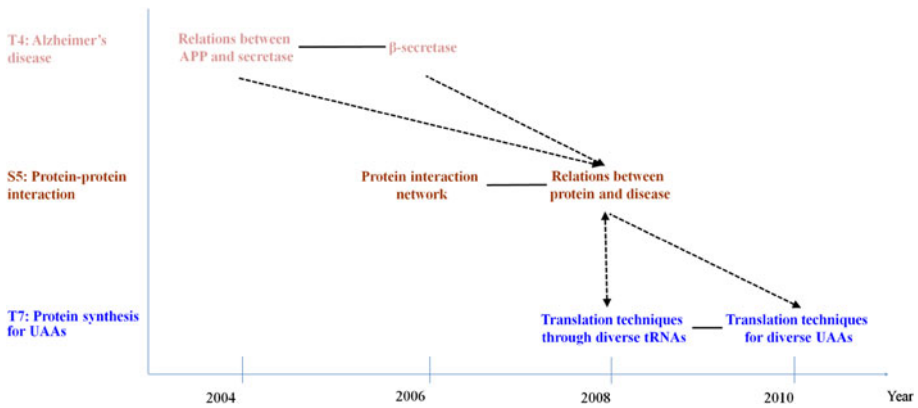
- **Maize breeding (T5):** This is the science of changing the genetics of plants to produce desired characteristics. Maize is a cross-pollinated crop, so later generations have distinct characteristics. Both self- and cross-pollination techniques were utilized in maize in 2006. In order to maintain the desired characteristics of variety, the maize hybrid variety is forcibly changed to inbred maize by manual control. In addition, later generations can possess many desired characteristics from multi-trait selection in hybrids with many stages. Thus, with multi-stage testing and multi-progeny evaluation, a number of patterns in these varieties with desired characteristics, such as high yield and quality in inbred or hybrid maize, were listed and published in 2008.
- **Diversity of seeds (T6):** With tissue culture technology and the inbred and hybrid genetic complements of plants, more corn varieties are being produced by crossing corn plants with themselves or with other plants of the same variety, which are then patented for commercialization. The conception of the right of new seed varieties became more prevalent in 2008. What's more, seed companies sell their seeds under their own names with patents, enjoying protection from utilization of their seeds by other producers. In 2010, the emergence of large companies led to market centralization, and the trend of oligopolistic competition became more obvious.
- **Protein synthesis for unnatural amino acids (UAAs) (T7):** This is a technology that translates UAAs into proteins. In 2008, the translation technology was developed through the orthogonal tRNAs, orthogonal aminoacyl-tRNA synthetases, or orthogonal pairs of tRNAs/synthetases in vitro translation. After that, the technology for the translation of diverse UAAs, such as *N*-acetylgalactosamine, heavy atom-containment, and keto amino acids were discussed in 2010. The changes of the structure and additional functions of proteins are significant because goal proteins and translated UAAs are critical components in pharmaceuticals and developmental drugs.
- **Taste receptors (T8):** The receptors of the sweet, umami and bitter tastes are composed of proteins of the family of G protein-coupled receptors (GPCR), such as the TAS1R, TAS2R and TAS3R proteins. Many studies focused on the function of receptors in 2008. After binding to a wide variety of substitutes, such as sugar and monosodium glutamate (MSG), the receptors send their signals to the brain via nerves related to the gustatory system. In 2010, in order to improve foods for commercialization according to these discoveries, the analysis of the combination of the test was used to search for or make molecules and substitutes to which taste receptors are sensitive.

#### Correspondence relations between scientific and technical fronts

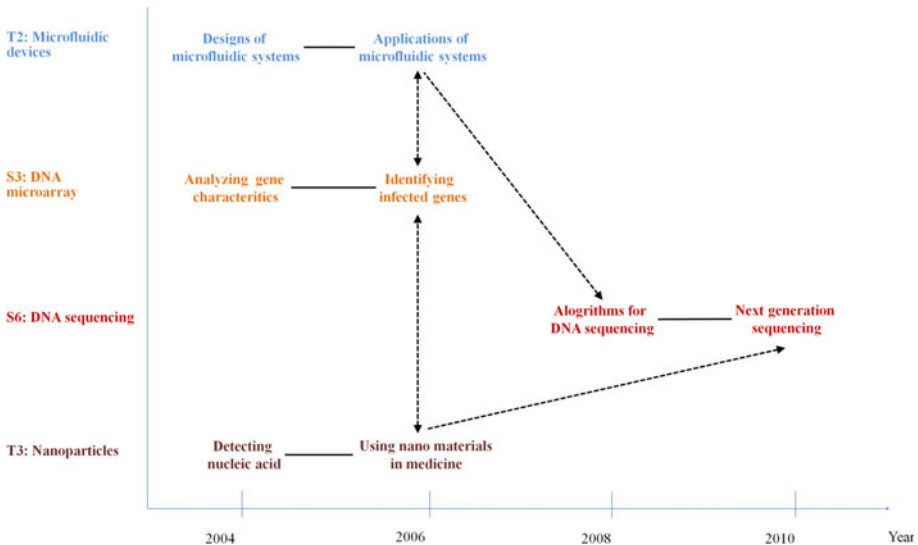
The following steps are carried out to check whether the scientific fronts and technical fronts exhibit lexical relations. First, the term vector of each front is established. Then, the CS was employed to calculate the similarity of each front pair between a science and a technology. The results of the CS between the fronts are presented in Table 1. To select front pairs with strong similarities, a front matching threshold of 0.138 was decided by Rosin thresholding. In sum, there are a total of 8 CSs above the threshold of 0.138, which were boldfaced in Table 1. These 8 CSs illustrate that there are 8 topics in both S&T in different and similar time periods. Since there may be a time difference, this paves the way for probing the temporal gap between them.



The trajectory topic “disease treatment” includes three matched trajectories: Alzheimer’s disease (T4), protein–protein interaction (S5), and protein synthesis for UAAs (T7), so they are discussed together. As shown in Fig. 4, both publications and patents focus on disease and protein interaction. In 2004, a patent study discussed the interaction between Alzheimer’s disease and protein. Then in 2006, patents for the application of protease inhibitors for the treatment of Alzheimer’s disease were granted. Because there is no cure for the disease, which worsens as it progresses, and eventually leads to death. It causes hundreds of thousands of elderly deaths each year. Currently there is no drug that can cure Alzheimer’s disease, but pharmaceutical companies have developed some drugs to slow its progression. On the other hand, in publication set, it focuses on the mechanisms of Alzheimer’s disease, trying to find new treatments to fight the disease. In 2006, publications used the concept of molecular biology and mathematical statistics to build a protein



**Fig. 4** Temporal gaps of fronts between trajectories (subjects related to disease treatment)



**Fig. 5** Temporal gaps of fronts between trajectories (subjects related to gene analysis platform)

interaction network system and protein–protein interaction. In 2008, human proteomic map was applied to disease diagnosis, treatment, and mechanisms. In later patent stages, it controls protein synthesis technology. Therefore, specific introduction of amino acid technology helps the study of protein function and structure, as well as protein mutation study.

The trajectory topic “gene analysis platform” includes four matched trajectories: microfluidic devices (T2), DNA microarray (S3), nanoparticles (T3), and DNA sequencing (S6), so they are discussed together. As shown in Fig. 5, both publications and patents involve in-depth research into genetic analysis. From 2004 through 2006, there were many patents on the development of microfluidic systems development and application of nano-materials, making detection of nucleic acid technology to be enhanced and new instruments were available for medical testing and research. These patent technologies and instruments created progress in bioinformatics of publications, such as the early 2004–2006 genetic analysis platform and the development of microarray technology, as well as the late 2008–2010 evolution of next-generation DNA sequencing. This evolution not only applies to rapid analysis of DNA samples, but also to the analysis of human genome variation, a specific drug design consideration.

## Conclusions and further suggestions

A series of procedures is performed to compare the front trajectories of S&T, in order to understand the relationship between S&T and the temporal gap between them in the biotechnology field. Researchers, government officials, or enterprises are able to gain understanding of worldwide S&T trends. Medical study and biotechnology are the two major subjects that participate in publication development, while patent development focuses on: plant breeding, disease treatment, and medical equipment improvement. We found that several scientific front trajectories are in the basic science stage, including embryonic or mesenchymal stem cells, RNA interference, microRNA, and microbial fuel cells, in which corresponding technology fronts are seldom found. On the other hand, several technical front trajectories have been practically applied, receiving less inspiration from science, such as plant breeding, seed diversity and taste receptors. Furthermore, although it is widely accepted that basic research in science provides a fundamental basis for technology-oriented innovations, this study shows that such a linear model is criticized in the biotechnology field, where scientific outputs also tend to be connected to technological inventions. We found that technology preceding science, synchronous development, and science preceding technology show connections between S&T in disease treatment and the gene analysis platform. In disease treatment, the technology of “secretase” preceded science into “relations between protein and disease”, and the latter science developed synchronously with and preceded the technology of “translation techniques”. In disease treatment, the technologies of “microfluidic systems” and “nano materials” developed synchronously with science into “identifying infected genes”, which then preceded science into “DNA sequencing”.

With the S&T connected subjects available, our ongoing studies provide potential opportunities for bridging publication affiliations and patent assignees. Our efforts are evidently shown in the flows of knowledge between affiliations and assignees, which lead to considerable diffusion of scientific and technical knowledge. It could be used also explore the influence on success of scientist-technologists in order to make effective policies for conflict of interest.

**Acknowledgments** This research is partially supported by the National Science Council, Taiwan, under contract No. NSC101-2811-B-002-003-.

## References

- Arber, W. (2009). The impact of science and technology on the civilization. *Biotechnology Advances*, 27(6), 940–944.
- Bassecoulard, E., & Zitt, M. (2005). Patents and publications. The lexical connection. In H. F. Moed, W. Glänzel, & U. Schmoch (Eds.), *Handbook of quantitative science and technology research* (pp. 665–694). Dordrecht: Kluwer.
- Boyack, K. W., Newman, D., Duhon, R. J., Klavans, R., Patek, M., Biberstine, J. R., et al. (2011). Clustering more than two million biomedical publications: Comparing the accuracies of nine text-based similarity approaches. *PLoS ONE*, 6(3), e18029.
- Braun, R. (2002). People's concerns about biotechnology: Some problems and some solutions. *Journal of Biotechnology*, 98(1), 3–8.
- Callaert, J., Grouwels, J., & Van Looy, B. (2012). Delineating the scientific footprint in technology: Identifying scientific publications within non-patent references. *Scientometrics*, 91(2), 383–398.
- Chen, S. H., Huang, M. H., & Chen, D. Z. (2012a). Identifying and visualizing technology evolution: A case study of smart grid technology. *Technological Forecasting and Social Change*, 79(6), 1099–1110.
- Chen, S. H., Huang, M. H., & Chen, D. Z. (2013). Exploring technology evolution and transition characteristics of leading countries: a case of fuel cell field. *Advanced Engineering Informatics*. doi: [10.1016/j.aei.2013.02.001](https://doi.org/10.1016/j.aei.2013.02.001).
- Chen, S. H., Huang, M. H., Chen, D. Z., & Lin, S. Z. (2012b). Detecting the temporal gaps of technology fronts: A case study of smart grid field. *Technological Forecasting and Social Change*, 79(9), 1705–1719.
- Engelsman, E. C., & van Raan, A. F. J. (1994). A patent based cartography of technology. *Research Policy*, 23(1), 1–26.
- Falkowski, T. (2009). *Community analysis in dynamic social networks*. Dissertation, University Magdeburg.
- Gardner, P. L. (1999). The representation of science-technology relationships in Canadian physics textbooks. *International Journal of Science Education*, 21(3), 329–347.
- Gavrilescu, M., & Chisti, Y. (2005). Biotechnology: A sustainable alternative for chemical industry. *Biotechnology Advances*, 23(7–8), 471–499.
- Glänzel, W., & Thijs, B. (2011). Using 'core documents' for the representation of clusters and topics. *Scientometrics*, 88(1), 297–309.
- Glänzel, W., & Czerwon, H. J. (1996). A new methodological approach to bibliographic coupling and its application to the national, regional and institutional level. *Scientometrics*, 37(2), 195–221.
- Glänzel, W., & Meyer, M. (2003). Patents cited in the scientific literature: An exploratory study of reverse citation relations. *Scientometrics*, 58(2), 415–428.
- Hall, B., Jaffe, A., & Trajtenberg, M. (2001). *The NBER patent citations data file: Lessons, insights and methodological tools*. WP 8498 National Bureau of Economic Research.
- Huang, Y. H. (2009). An investigation of the interaction between science and technology from literatures of giant magnetoresistance. *Journal of Library and Information Studies*, 7(1/2), 53–76.
- Kostoff, R. N., & Schaller, R. R. (2001). Science and technology roadmaps. *IEEE Transaction on Engineering Management*, 48(2), 132–143.
- Lee, C. H., & Chien, T. F. (2013). Leveraging microblogging big data with a modified density-based clustering approach for event awareness and topic ranking. *Journal of Information Science*. doi: [10.1177/0165551513478738](https://doi.org/10.1177/0165551513478738).
- Lissoni, F., & Montobbio, F. (2008). *Inventorship and authorship in patent-publication pairs: An enquiry into the economics of scientific credit*, Working Papers 224. Centre for Knowledge, Internationalization and Technology Studies, Università Bocconi, Milano.
- Lo, S. S. (2010a). Scientific linkage of science research and technology development: A case of genetic engineering research. *Scientometrics*, 82(1), 109–120.
- Lo, S. S. (2010b). A comparative study of linkage indexes: Co-assignee, reciprocal citation, patent coupling and co-patent. *Journal of Library and Information Studies*, 8(1), 11–27.
- Nakamura, H., Sasaki, H., Shibata, N., Kajikawa, Y., Sakata, I., & Suzuki, S. (2010). Science and technology map analysis of a multi-disciplinary field-case study of aerospace engineering. In *IEEE international conference on industrial engineer and engineering management*, (pp 1901–1905).
- Narin, F., Hamilton, K., & Olivastro, D. (1997). The increasing linkage between US technology and public science. *Research Policy*, 26(3), 317–330.

- Narin, F., & Olivastro, D. (1998). Linkage between patents and papers: An interim EPO/US comparison. *Scientometrics*, 41(1–2), 51–59.
- Newman, M. E. J. (2004). Analysis of weighted networks. *Physical Review E*, 70(5), 056131.
- Perng, D. B., & Chen, S. H. (2011). Directional textures auto-inspection using discrete cosine transform. *International Journal of Production Research*, 49(23), 7171–7187.
- Ribeiro, L. C., Ruiz, R. M., Bernardes, A. T., & Albuquerque, E. M. (2010). Matrices of science and technology interactions and patterns of structured growth: Implications for development. *Scientometrics*, 83(1), 55–75.
- Salton, G., & McGill, J. M. (1983). *Introduction to modern information retrieval*. New York: McGraw-Hill.
- Shalhevet, S., Haruvy, N., & Spharim, I. (2001). Management strategies for agricultural biotechnology in small countries: A case study of Israel. *Biotechnology Advances*, 19(7), 539–554.
- Shibata, N., Kajikawa, Y., & Sakata, I. (2011). Detecting potential technological fronts by comparing scientific papers and patents. *Foresight*, 13(5), 51–60.
- Small, H. (2006). Tracking and predicting growth areas in science. *Scientometrics*, 68(3), 595–610.
- Small, H., & Griffith, B. (1974). The structure of scientific literatures, I: Identifying and graphing specialities. *Science Studies*, 4, 17–40.
- Verbeek, A., Debackere, K., Luwel, M., Andries, P., Zimmermann, E., & Deleus, F. (2002). Linking science to technology: Using bibliographic references in patents to build linkage schemes. *Scientometrics*, 54(3), 399–420.
- Wu, H. C., Chen, H. Y., & Lee, K. Y. (2010). Unveiling the core technology structure for companies through patent information. *Technological Forecasting and Social Change*, 77(7), 1167–1178.