

A bibliometric analysis of gene editing research^①

Yuan Fang (袁芳)^②, Sun Jingsong, Zheng Jia, Li Nong

(Institute of Scientific and Technical Information of China, Beijing 100038, P. R. China)

Abstract

Gene editing, as a breakthrough in life science and medicine, makes significant scientific impact. The entire development process of gene editing through bibliometric analysis is studied to explore the distribution features and research focus with visual tool. The number of annual publications distribution, journal distribution and country distribution are analyzed in the basic distribution analysis. The keywords and co-citation are analyzed with CiteSpace to explore the research hotspots. The number of publications related to gene editing is increasing year by year. USA is the main publishing countries for gene editing research. CRISPR/Cas system is a hot topic in the field of genetic editing in last few years. Gene editing is opening up a world of possibilities for the treatment of genetic diseases.

Key words: gene editing, bibliometric analysis, CiteSpace, publication

0 Introduction

Gene editing is a type of genetic engineering in which DNA is inserted, deleted, modified or replaced in the genome of a living organism. Because of gene editing targeting the insertions to site specific locations, gene editing has precision and efficiency features. The precise gene editing is a powerful tool for studying biological processes^[1]. Gene editing is widely used to create model cell lines, engineer metabolic pathways, produce transgenic animals and plants, perform genome-wide functional screen and, most importantly, treat human diseases that are difficult to tackle by traditional medications^[2]. The major hurdles for efficient gene editing in therapeutically relevant primary human cells have largely been overcome as shown in numerous recent preclinical studies and multiple clinical trials are expected to start within the next couple of years^[3]. In recent years, gene editing technology has been selected as “disruptive technology” by many organizations. Gene editing has taken the biomedical science field by storm, initiating rumors about future Nobel Prizes and heating up a fierce patent war, but also making significant scientific impact^[4].

The common methods for gene editing used engineered nucleases, and 4 major families of engineered nucleases were used: meganucleases, zinc finger nucleases (ZFNs), transcription activator-like effector

based nucleases (TALENs), and the clustered regularly interspaced short palindromic repeats (CRISPR/Cas9) system^[5-8]. Meganucleases, discovered in the late 1980s, are enzymes in the endonuclease family which are characterized by their capacity to recognize and cut large DNA sequences (from 14 to 40 base pairs)^[9]. ZFNs and TALENs are chimeric Zinc fingers and TAL effectors are fused to the endonucleolytic DNA cleavage domain of the FokI endonuclease to make zinc-finger proteins or transcription activator-like effector proteins^[2]. CRISPR/Cas9 system relies on single guide RNA directing the Cas9 endonuclease to a specific site in the genome^[3]. Due to different features of the 4 approaches, meganucleases, ZFNs, TALENs and CRISPR/Cas9 system have their applicable fields. TALENs and engineered meganucleases were selected by *Nature Methods* as the 2011 Method of the Year^[10]. The CRISPR/Cas9 system was selected by *Science* as 2015 Breakthrough of the Year^[11].

Bibliometric analysis is a basic but an effective way to detect and examine the emergence of a new technology^[12]. At present, Cao et al.^[13], Liu^[14] and Wang et al.^[15] studied the field of gene editing through bibliometrics. However, these researchers just studied the recent 10 years development of gene editing. This research studies the entire development process of gene editing through bibliometric analysis to explore the distribution features and research focus with visual tool.

① Supported by China Postdoctoral Science Foundation (No. 2017M610970).

② To whom correspondence should be addressed. E-mail: yuanf@istic.ac.cn
Received on July 9, 2019

1 Method and data collections

The bibliometric study involves the statistical analysis of scientific publications, which adopts quantitative performance indicators to get over the disadvantage of subjectivity in peer review and expert judgments, and has been used to assess research performance in an increasing amount and variety of studies^[16]. These have laid the groundwork for implementing the basic scientometric analysis which evaluates emerging research domains in the present study^[17]. Visual analysis of scientific publications is an important branch of information visualization. CiteSpace is a Java-based scientific visualization tool, which could be potentially used to detect and visualize abrupt changes, emerging trends, and dynamics in scientific domains^[18]. In this paper, we will use CiteSpace to describe the development process of gene editing.

The data were retrieved from Science Citation Index Expanded (SCI-EXPANDED) of the Web of Science Core Collection database. Search string is TS (Topic) = (“gene editing” OR “genome editing” OR “gene edited” OR “genome edited” OR “gene editMYM” OR “genome editMYM” OR (“clustered regularly interspaced short palindromic repeats”) OR (“clustered regularly interspaced short palindromic repeat”) OR (“CRISPRMYM”) OR (“CRISPER” and “cas”) OR (“CRISPER” and “cas9”) OR (“meganuclease” OR “meganucleases” OR “homing endonuclease \$”) OR (((“ZFN” OR “ZFNS”) and (“gene” or “genome” or “nuclease \$”))) OR (“zinc fingers”) OR (“zinc finger”)) OR ((“transcription activator-like effector \$”) OR (“TAL effector \$”) OR ((“TALEN” OR “TALENS”) and (“gene” OR “genome” OR “nuclease \$”))) OR (“transcriptional activator-like effector \$”))). Document type set as “ARTICLE”. The selected results were sent to file of plain text by setting the “record content” as “full record and cited references”. In basic distribution analysis, the database chosen for gene editing obtained 31 157 articles as of 2018. In keywords analysis and co-citation analysis, the database chosen for this study obtained 32 625 articles and data set last updated on April 19, 2019. CiteSpace was used in keywords analysis and co-citation analysis. After the data cleaning and the data preprocessing through CiteSpace, keywords screening includes artificial analysis. In artificial analysis, some words that are not related to gene editing technology or have no practical meaning will be eliminated. For example, we eliminated these words like “family”, “specificity” and

“nf kappa b”, etc.

2 Results and analysis

2.1 Basic distribution analysis

2.1.1 The number of annual publications distribution analysis

The number of annual publications is an important index for measuring the development of scientific research, as it reflects, to a certain extent, the changes in knowledge quantity^[19]. Therefore, the distribution of annual publications (Fig. 1) can form a preliminary understanding of the development process of gene editing. Because the publications data is incomplete in 2019, we only count it as of 2018. It can be seen that the first gene editing related article was published in 1987 and 31 157 gene editing related articles had been published from 1987 to 2018.

In 1996, ZFN was constructed by Kim et al.^[20]. One major concern of using ZFNs for clinical applications is its potential off-target activity, for instance, to mutate both alleles of the dihydrofolate reductase gene in Chinese hamster ovary cells to improve production of recombinant protein or to create human blood cells resistant to HIV by mutating the gene encoding CCR5^[21,22]. In 2010, TALEN was first constructed by Christian^[23]. TALENs are used in a similar way to ZFNs. In practice, ZFNs and TALENs require a lot of time and cost to build and design manually. Since the Cas9-sgRNA system was developed as a genome editing system in 2012^[24], the number of publications related to gene editing has grown rapidly with hundreds of related articles published in every year. In 2018, 4 284 gene editing related articles was published, which is the most ever for a single year from 1987 to 2018. CRISPR/Cas has the advantages of simple operation and low cost. However, the CRISPR/Cas system's editable genomic loci are limited by DNA sequences, and the off-target effect also limits the application of CRISPR/Cas system. Therefore, meganucleases, ZFN and TALEN with their own unique technical advantages have not been completely replaced by the CRISPR/Cas system. Gene editing technology, especially CRISPR/Cas, has been rated as one of the most important technological breakthroughs in the field of life sciences. Significant advances in gene-editing technology are emerging, bringing disruptive changes to life science research and related industries. Governments all over the world attach great importance to the application and development of gene editing and have invested a lot of funds to support the research of gene editing technology.

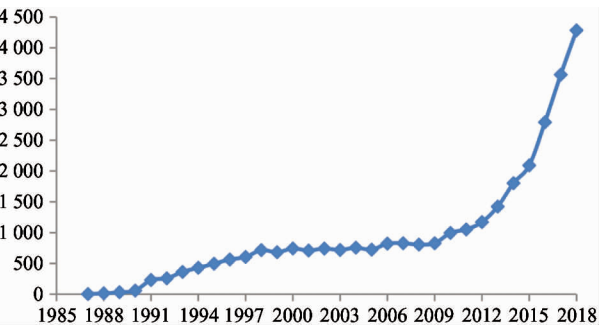


Fig. 1 Yearly publications of gene editing research output from 1987 to 2018

2.1.2 Journal distribution analysis

Journal analysis provides guidance for scholars to select platforms for data collection and publishing their researches^[19]. 31 157 gene editing related articles published from 1987 to 2018 and the top 15 academic journals are shown in Table 1. “Journal of Biological Chemistry”, “PloS One” and “Nucleic Acids Research” have issued more than 1 000 gene editing related articles. “Journal of Biological Chemistry” tops the list with 1 477 publications, manifesting the journal

greatly favors the subject of gene editing. The impact factors of “Nucleic Acids Research”, “Nature Communication” and “EMBO Journal” all are greater than 10, which proves there is a high possibility of publishing research articles on gene editing in high-level journals.

In addition, the categories of “PloS One”, “Proceedings of the National Academy of Sciences”, “Scientific Reports” and “Nature Communication” all belong to multidisciplinary sciences, which shows gene editing related articles are not only issued on biological journals but also on multidisciplinary journals that would favor the gene editing researches. Gene editing technology, as one of the most important breakthroughs in biotechnology in this century, is gradually intermingling with disciplines such as chemistry, information, and materials. It has conceived and promoted a number of new technologies and new applications in medicine, agriculture, industry, etc. This may be why gene editing technology is widely published in these journals of multidisciplinary sciences.

Table 1 The top 15 journals in the field of gene editing

Rank	Journal title	Count	Value(%)	IF2018
1	Journal of Biological Chemistry	1 477	4. 741	4. 106
2	PloS One	1 037	3. 328	2. 776
3	Nucleic Acids Research	1 008	3. 235	11. 147
4	Proceedings of the National Academy of Sciences	926	2. 972	9. 580
5	Scientific Reports	775	2. 487	4. 011
6	Molecular and Cellular Biology	598	1. 919	3. 735
7	Biochemical and Biophysical Research Communications	553	1. 775	2. 705
8	Development	480	1. 541	5. 763
9	Journal of Virology	364	1. 168	4. 324
10	Nature Communication	340	1. 091	11. 878
11	Oncogene	332	1. 066	6. 634
12	Developmental Biology	325	1. 043	2. 936
13	EMBO Journal	288	0. 924	11. 227
14	Gene	287	0. 921	2. 638
15	Journal of Molecular Biology	283	0. 908	5. 067

2.1.3 Country distribution analysis

In the field of gene editing, a total of 144 countries or regions have published relevant articles as of 2018 and Fig. 2 shows the top 15 countries or regions. The number of gene editing related articles in USA is 14 525, accounting for 46. 619% of the world total, which is far ahead of the rest of the world. It indicates that USA is the main publishing countries for gene editing research. The number of China’s gene editing related articles is followed by 5 237, accounting for

16. 808% of the world total. Japan ranks the 3rd in the world, and the number of Japan’s gene editing related articles is 3 228, accounting for 10. 360% of the world total. Other countries’ gene editing related articles account for less than 10% of the world total.

In USA, some institutes such as the Massachusetts Institute of Technology, Harvard University, and the Broad Institute have done a lot of outstanding works in the field of gene editing technology. In addition, many companies also promote the development of gene edi-

ting technology. For example, Sangamo develops clinical medicine for gene editing technology and Corteva develops gene editing agricultural products. In China, the Chinese Academy of Sciences has the most outstanding scientific research achievements in the field of gene editing technology. Comparing with USA, Chinese companies engaged in gene editing technology are relatively weak.

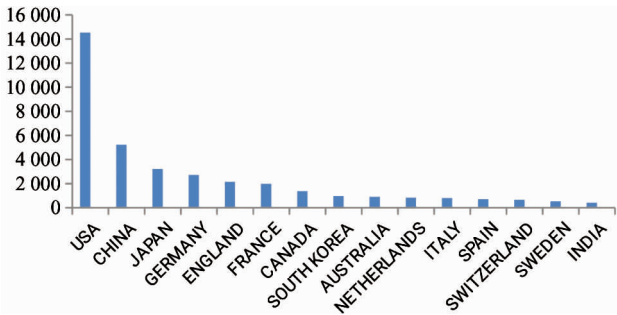


Fig. 2 Countries or regions of gene editing researches in the world

2.2 Keywords analysis

2.2.1 The time-zone of keywords analysis

By plotting the time zone of keywords, the dynamic development process of the gene editing research can be obtained Fig. 3. Gene editing research originated in the late 1980s and early 1990s. The early gene editing research mainly focused on the protein level research of nucleases. In this period, some keywords emerged, such as “protein”, “expression”, “crystal structure”. In the middle stage of gene editing development, the research mainly focused on the mechanism level of nucleases. “Apoptosis”, “phosphorylation”, “Mechanism”, “pathway” and so on are merged in this period. After 2007, “zinc finger nuclease” as the representative of gene editing nuclease has emerged on the research of gene editing. And CRISPR has been the focus of gene-editing research in last five years. In recent years, gene editing has also been applied to human cells to treat the diseases. “stem cell”, “disease”, “human cell” and “therapy” are merged in this period.

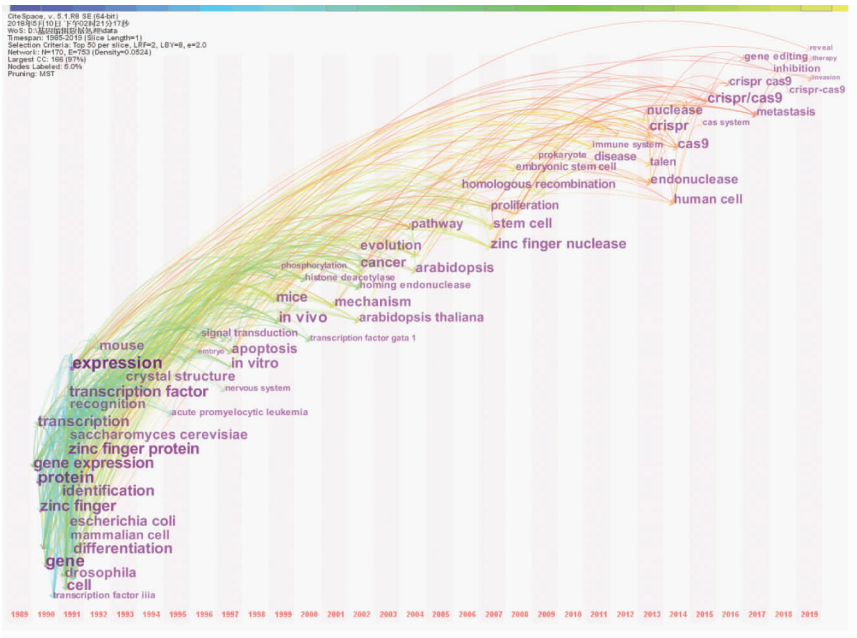
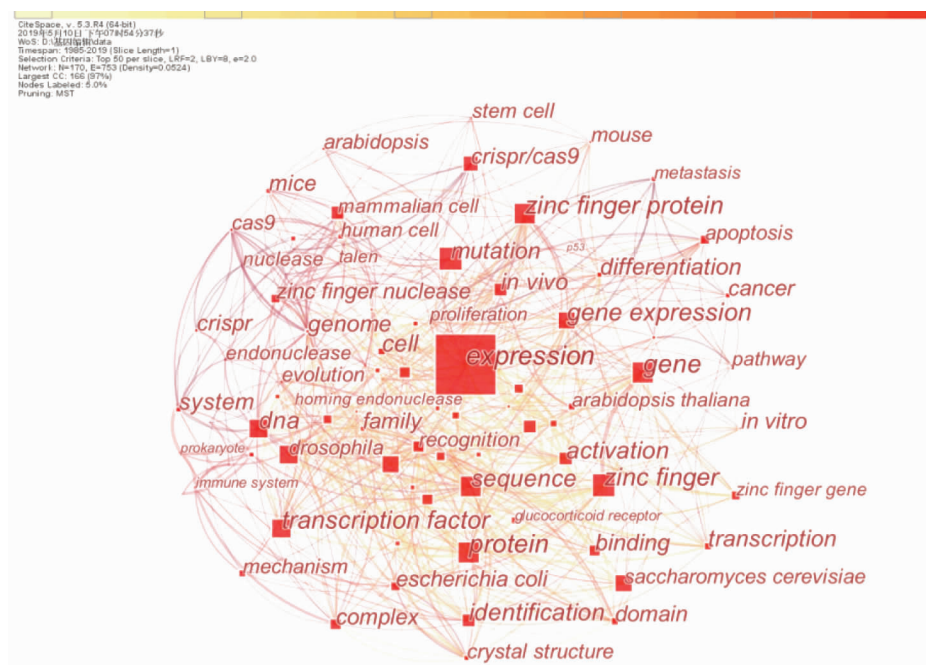


Fig. 3 The time-zone of keywords in the field of gene editing

2.2.2 Co-occurrence network of keywords analysis

Keywords indicate the soul of articles and can show at a glance the advancement in research topics. Co-occurrence network of keywords analysis can evaluate the topics and research trends in the field of gene editing. In Fig. 4, the network has 170 nodes indicating the number of keywords and 1 332 links. The size of the node represents the citation count of the keyword, and the larger the node is, the higher the cita-

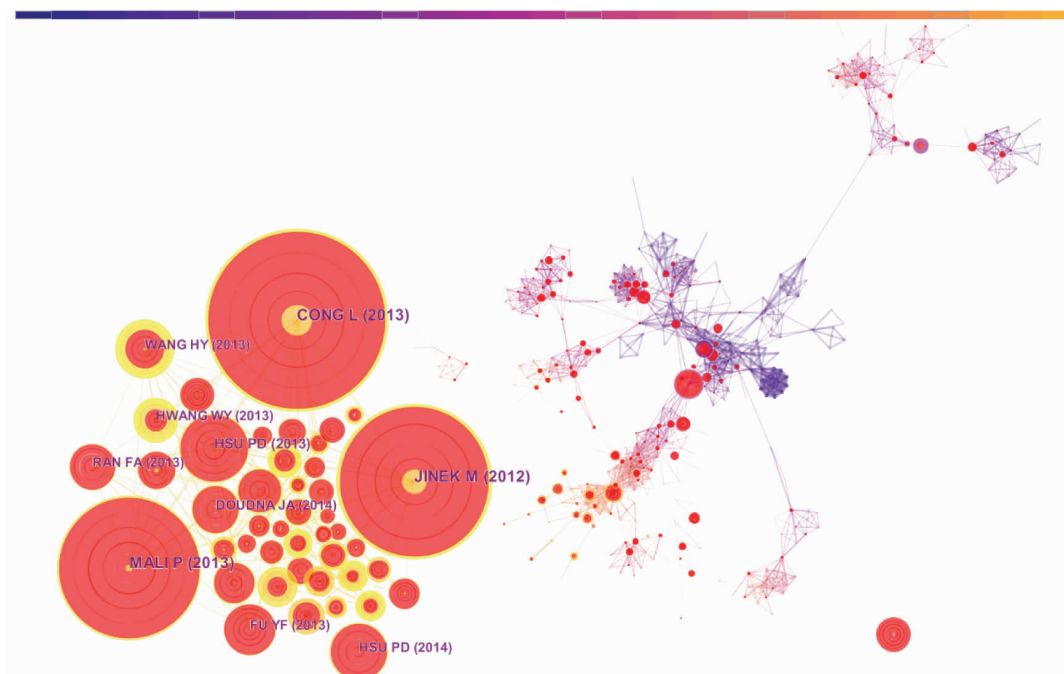
tion count is. The citation counts of “expression”, “gene”, “protein”, “zinc finger protein” and “gene expression” are all above 3 000. The links between the nodes represent the co-occurrence relationship of the keywords. The links between “zinc finger protein” and “gene expression” are dense, which shows the relationship of these 2 keywords is close that these 2 keywords are likely to appear in the same literature.



2.3 Co-citation analysis

Co-citation analysis is one key aspect of bibliometrics, which can identify the influence of the authors and the literatures^[25]. The intellectual structure of highly cited literatures can be developed by means of CiteSpace. The “Top N per slice” is set as $N = 50$, that means 50 most cited or occurred items are selected from each slice. Fig.5 shows the salient intellectual structure of co-cited literatures. In Fig.5, Some points

on the right are not connected with points on the left, indicating that in the first 50 documents of each year, they are not co-cited with these literatures in the figure. Most of the points on the left are highly cited literatures, which concentrate in CRISPR/Cas9. For example, Jinek^[24] et al. first demonstrated CRISPR/Cas9 could be used in gene editing system. Cong et al.^[26] and Mali et al.^[27] respectively used CRISPR/Cas 9 in human cell lines, which has greatly promoted



the application of CRISPR technology. The points on the left are some lower cited literatures, which include the relatively little research in gene editing technology, such as meganucleases.

From the network summary table exported by CiteSpace, the top 10 most cited literatures in the co-citation network are selected for specific analysis, as shown in Table 2. The researches of top 10 highly cited literatures are all about CRISPR/Cas and concentrated in 2013. Due to the high concentration of re-

search content and publication time, some literatures did not have a co-citation relationship with the top 10 highly cited literatures in Fig. 5. This also explains that CRISPR/Cas is a hot topic in the field of genetic editing. Compared with other methods of gene editing, CRISPR/Cas9 has the advantages of simple operation and low cost. Hence, CRISPR/Cas9 has developed rapidly since 2012 and the highly cited literatures concentrated in CRISPR/Cas.

Table 2 The top 10 highly cited literatures in the field of gene editing

Rank	Title	First author	Year	Citations
1	Multiplex genome engineering using CRISPR/Cas systems	Cong L	2013	2 368
2	A programmable Dual-RNA-Guided DNA endonuclease in adaptive bacterial immunity	Jinek M	2012	2 026
3	RNA-Guided human genome engineering via Cas9	Mali P	2013	1 882
4	Genome engineering using the CRISPR-Cas9 system	Ran F A	2013	1 013
5	DNA targeting specificity of RNA-guided Cas9 nucleases	Hsu P D	2013	915
6	One-Step generation of mice carrying mutations in multiple genes by CRISPR/Cas-Mediated genome engineering	Wang H Y	2013	779
7	Development and applications of CRISPR-Cas9 for genome engineering	Hsu P D	2014	768
8	High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells	Fu Y F	2013	688
9	Structures of Cas9 endonucleases reveal RNA-Mediated conformational activation	Doudna J A	2014	637
10	Efficient genome editing in zebrafish using a CRISPR-Cas system	Hwang W Y	2013	622

3 Conclusions

In this study, the basic distribution of gene editing is analyzed, including the number of annual publications distribution, journal distribution and country distribution. The number of publications related to gene editing is increasing year by year. Since 2011, the number of publications has grown rapidly with hundreds of related articles published in every year. USA is the main publishing countries for gene editing research, which number of gene editing related articles is almost half of the world. Gene editing, as “disruptive technology”, has taken the biomedical science field by storm. Therefore, both of high-level journals or multidisciplinary journals would favor and publish the articles of gene editing research. In order to further study, the research hotspots are explored with CiteSpace to analysis the keywords and co-citation. In the 4 major methods of gene editing, the researches of ZFNs and CRISPR/Cas9 system are relatively more than those of meganucleases and TALEN. CRISPR/Cas system is a hot topic in the field of genetic editing in last few

years. Gene editing is opening up a world of possibilities for the treatment of genetic diseases.

Reference

[1] Mcmahon M A, Rahdar M, Porteus M. Gene editing: not just for translation anymore[J]. *Nature Methods*, 2011, 9(1): 28-31

[2] Liu J, Shui S L. Delivery methods for site-specific nucleases: achieving the full potential of therapeutic gene editing[J]. *Journal of Controlled Release*, 2016, 244 (Pt A): 83-97

[3] Bak R O, Natalia G O, Porteus M H. Gene editing on center stage[J]. *Trends in Genetics*, 2018, 34(8): 600-611

[4] Blenke E O, Evers M J W, Mastrobattista E, et al. CRISPR-Cas9 gene editing: delivery aspects and therapeutic potential[J]. *Journal of Controlled Release*, 2016, 244(Pt B): 139-148

[5] Osborn M J, Belanto J J, Tolar J, et al. Gene editing and its application for hematological diseases[J]. *International Journal of Hematology*, 2016, 104(1): 18-28

[6] Esvelt K M , Wang H H. Genome-scale engineering for systems and synthetic biology[J]. *Molecular Systems Biology*, 2014, 9(1):641-641

[7] Tan W S, Carlson D F, Walton M W, et al. Precision

- editing of large animal genomes [J]. *Advances in Genetics*, 2012, 80: 37-97
- [8] Puchta H, Fauser F. Gene targeting in plants: 25 years later[J]. *The International Journal of Developmental Biology*, 2013, 57(6-8): 629-637
- [9] Stoddard B L. Homing endonuclease structure and function[J]. *Quarterly Reviews of Biophysics*, 2005, 38(1): 49-95
- [10] None. Method of the year 2011 [J]. *Nature Methods*, 2011, 9(1): 1-1
- [11] Travis J. Breakthrough of the year: CRISPR makes the cut [EB/OL]. <https://www.sciencemag.org/news/2015/12/and-science-s-2015-breakthrough-year>; Sciencemag, 2015
- [12] Braun T, Schubert A. A quantitative view on the coming of age of Interdisciplinarity in the sciences, 1980-1999 [J]. *Scientometrics*, 2003, 58(1): 183-189
- [13] Cao X W, Gao X S, Chen R. Analysis of international development trend of gene editing technology based on bibliometrics[J]. *Global Science, Technology and Economy Outlook*. 2018, 33(4): 60-72. (In Chinese)
- [14] Liu Q. Analysis of the development of genome editing technology based on literature[J]. *Science and Technology & Innovation*, 2019, (10): 37-39. (In Chinese)
- [15] Wang H Y, Yuan T W, Ruan L L, et al. Development trend analysis of genome editing technologies based on bibliometric[J]. *Biotechnology & Business*, 2017, (1): 13-20. (In Chinese)
- [16] Rinia E J, van Leeuwen T N, van Vuren H G, et al. Comparative analysis of a set of bibliometric indicators and central peer review criteria: evaluation of condensed matter physics in the Netherlands[J]. *Research Policy*, 1998, 27(1): 95-107
- [17] Chen K, Guan J. A bibliometric investigation of research performance in emerging nanobiopharmaceuticals [J]. *Journal of Informetrics*, 2011, 5(2): 233-247
- [18] Chen C. CiteSpace II: detecting and visualizing emerging trends and transient patterns in scientific literature[J]. *Journal of the American Society for Information Science and Technology*, 2006, 57(3): 359-377
- [19] Qi T, Wang T, Ma Y, et al. A scientometric analysis of e-participation research [J]. *International Journal of Crowd Science*, 2018, 2(2): 136-148
- [20] Kim Y G, Cha J, Chandrasegaran S. Hybrid restriction enzymes: zinc finger fusions to FokI cleavage domain [J]. *Proceedings of the National Academy of Sciences*, 1996, 93(3): 1156-1160
- [21] Perez E E, Wang J, Miller J C, et al. Establishment of HIV-1 resistance in CD4 + T cells by genome editing using zinc-finger nucleases [J]. *Nature Biotechnology*, 2008, 26(7): 808-816
- [22] Holt N, Wang J, Kim K, et al. Human hematopoietic stem/progenitor cells modified by zinc-finger nucleases targeted to CCR5 control HIV-1 in vivo[J]. *Nature Biotechnology*, 2010, 28(8): 839-847
- [23] Christian M, Cermak T, Doyle E L, et al. Targeting DNA double-strand breaks with TAL effector nucleases [J]. *Genetics*, 2010, 186(2): 757-761
- [24] Jinek M, Chylinski K, Fonfara I, et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity[J]. *Science*, 2012, 337(6096): 816-821
- [25] Wang Q. Distribution features and intellectual structures of digital humanities [J]. *Journal of Documentation*, 2018, 74(1): 223-246
- [26] Cong L, Ran F A, Cox D, et al. Multiplex genome engineering using CRISPR/Cas systems[J]. *Science*, 2013, 339(6121): 819-823
- [27] Mali P, Yang L, Esvelt K M, et al. RNA-guided human genome engineering via Cas9 [J]. *Science*, 2013, 339(6121): 823-826

Yuan Fang, born in 1989. She works as a post-doctoral fellow in Institute of Scientific and Technical Information of China. She received her Ph. D degree in chemical genomics of Peking University in 2016. She also received her B. S. degree from Guangzhou University of Chinese Medicine in 2011. Her research interest is competitive intelligence research in biomedical industry.