

Visualization Analysis of CRISPR Gene-editing Knowledge Map based on Citespace

Can Gao^{a,b}, Rui Wang^b, Lin Zhang^{c,*}, and Changwu Yue^{a,**}

^a Yan'an Key Laboratory of Microbial Drug Innovation and Transformation, School of Basic Medicine, Yan'an University, Yan'an, Shaanxi, 716000 China

^b Non-coding RNA and Drug Discovery Key Laboratory of Sichuan Province, Chengdu Medical College, Chengdu, Sichuan, 610500 China

^c Department of Pharmacy, Shaoxing People's Hospital, Shaoxing Hospital, Zhejiang University School of Medicine, Shaoxing, Zhejiang, 312000 China

*e-mail: zhanglinfudan@zju.edu.cn

**e-mail: changwuyue@126.com

Received February 4, 2021; revised June 3, 2021; accepted June 21, 2021

Abstract—CRISPR is an adaptive immune defense system found in bacteria and archaea that is resistant to heterologous invasive genetic material. Later studies showed that the CRISPR system can be used for gene-editing. This study used the Web of Science database as a search object, then visually analyzed the literature related to CRISPR gene-editing technology with CiteSpace IV. The results show that publications had increased year by year. USA ranked first in terms of publications. China is second, but the centrality is very low. Doudna JA and Zhang F have made outstanding contributions. There are close connections between the internal institutions of the various states, but there are few links between the states. The hot spot and frontier are the application of CRISPR in animals, plants, detection, diagnosis, and clinical treatment.

Keywords: CRISPR, CiteSpace, Gene-editing, Visualization analysis

DOI: 10.1134/S1062359021060108

INTRODUCTION

Gene-editing is a tool for exploring gene function and important research content in the post-genome era. The original genome editing knocked out or replaced endogenous genes mainly through the HR (homologous reorganization) in the natural state. However, the efficiency of the traditional random HR is only one in a million, it is difficult to achieve fixed-point editing of the target gene (Keeney et al., 1997). At the beginning of the 20th century, researchers developed ZFN (zinc finger nuclease) and TALEN (Transcriptional activator-like effectors nuclease). But limited to several organisms (yeast, mouse, and human, etc.). Later in early 2012, a gene-editing technology that does not rely on FokI nuclease-CRISPR/Cas (CRISPR-associated, Cas) was reported.

CRISPR system itself which exists in bacteria and archaea genome is a kind of important genetic adaptive immune system (Cong et al., 2013). CRISPR sequence belongs to the family of DNA specific sequences, CRISPR loci are usually made of short length is 21–48 bp of highly conserved repeated sequences that are separated by non-repeated interval

sequences of 26–72 bp (Heler et al., 2014), CRISPR identifies the exogenous DNA sequence with these intervals. In 2015, Cell published an article called “SnapShot: CRISPR-RNA-Guided Adaptive Immune Systems”, which completely analyzed the “central rule” of the CRISPR system (Carter and Wiedenheft, 2015). The CRISPR system works in three steps. Stage 1: Foreign DNA acquisition, the host's Cas protein to identify foreign nucleic, and the short-segment DNA of the invading bacteria are called protospacers, which are inserted into the host's CRISPR as a spacer. Stage 2: CRISPR RNA (crRNA) biosynthesis, CRISPR DNA is transcribed to pre-crRNA, which is then processed by endonuclease into a library of short CRISPR-derived RNAs (crRNAs) contains a sequence that is complementary to the foreign DNA. Stage 3: Target Interference, mature crRNA directs Cas to bind to a complementary target such that the target sequence is degraded by a particular Cas. Currently, six types of CRISPR/Cas systems have been identified: class one system includes Types I, III, and IV. And class two system includes three types II, V, and VI. The most widely used Type II CRISPR/Cas requires only one Cas9 (Csn1/Cas5), the structure is relatively simple, which means that

CRISPR/Cas9 has greater application potential and development.

CRISPR's high editing efficiency is attracting research attention around the world. The uniqueness and efficiency of CRISPR have gradually aroused research attention around the world, and improvements in biological technologies have led to an avalanche-like accumulation of information. Therefore, creating and using web-based information solutions to navigate and process this kind of data is an urgent and necessary task for productive scientific work. CiteSpace is a free Java application for visualizing key concepts and trends based on the analysis of published data in a given subject area, which was invented by professor Chaomei Chen (Chen, 2006). Knowledge Graph will be applied mathematics, graphics, information visualization technology, theory and method of information science disciplines and metrology method such as citation analysis, co-occurrence analysis, and using the visualization of the image of the map to show discipline core structure, development history, the frontier, and the overall knowledge structure to achieve the purpose of multidisciplinary integration. This study aims to use CiteSpace to analyze the retrieved literature, construct a collaborative network in a visual way to analyze the research hotspots and development trends, as well as the relationships among countries, institutions, and authors, which provides new approaches and ideas for CRISPR gene-editing.

MATERIALS AND METHODS

Literature Retrieval

This study selected the core collection of the foreign language database Web of Science (WoS) to search for keywords. Due to the relatively broad concept of CRISPR, there were more subject areas and related topics. In order to make the retrieved literature highly versatile and inclusive, selected CRISPR gene-editing or engineering as keywords. The search path is (TS = (CRISPR*) AND TS = (edit* OR engineer*)), the language is English, document type: Article OR Review, index = SCI-EXPANDED, time span = 2009–2019 (retrieved date August 29, 2019). After refining and screening, a total of 6150 references were obtained. It's worth noting that the literature retrieved in 2019 is incomplete due to time constraints.

Statistical Methods

The data of literature included in this study were exported in the format of Refworks, the output file format of the rename is 'download_*.txt', references were imported into CiteSpace 4.0.r5 SE (64-bits) for data conversion. The parameters of CiteSpace were set as follows: The selection criteria is Top N: Top 100,

time slicing (1999–2019), years per slice (1), term source (all selection), node type (choose one at a time), selection criteria (top 50 objects) and pruning (pathfinder). The main procedural steps of CiteSpace software are time slicing, thresholding, modeling, pruning, merging, and mapping. The research is divided into three parts: (1) the visual analysis of literature mainly includes the number of publications and high citations analysis; (2) the analysis of knowledge subjects mainly includes authors, institutions, countries, and journals, as well as the analysis of cooperative graphs; (3) the research focus and frontier analysis mainly include the frequency, centrality, and clustering of keywords, the strength of mutation words and the analysis of time zone diagram. VOSviewer1. 6. 10 was used to select authors, institutions and conduct cooperative network analysis to draw the visualized graph.

In statistics, the two concepts of "frequency" and "centrality" are often involved. "Frequency" refers to the number of times a subject has occurred. "Centrality" can reflect the impact of research objects in the entire field, only by combining the two can we infer the general authority of the field of study. The greater the centrality, the higher representation of the corresponding research content in a certain period in the subject area. Betweenness centrality scores are normalized to the interval [0, 1]. Words with a central value >0.10 are the main keywords, indicating that the research object is meaningful. However, the centrality value of the cited literature is 1, which indicates that the cited are from the same reference and do not have significance. N represents the number of nodes, E represents the number of connections. The higher the density value is, the better the clustering result of the network is.

Besides, burst detection is performed on keywords, an algorithm for identifying sudden changes in events and other types of information. Keywords map types support two visualization views: cluster views and timezone views. Keywords timeline was used for research frontier analysis, the log-likelihood tests (LLR) is used to calculate the clustering label. If the clustering value is 1 and the number of members is less than 10, the cluster will not be displayed. Clustering with the same keywords indicates that has high attention at different time nodes.

RESULTS

Visual Analysis of the Literature

Analysis of Quantity and Growth Trend of Annual Publications. Time-related literature statistics can reflect the development of this subject. According to statistics from WoS, 6,150 articles were retrieved. As can be seen from Fig. 1, only one paper was produced

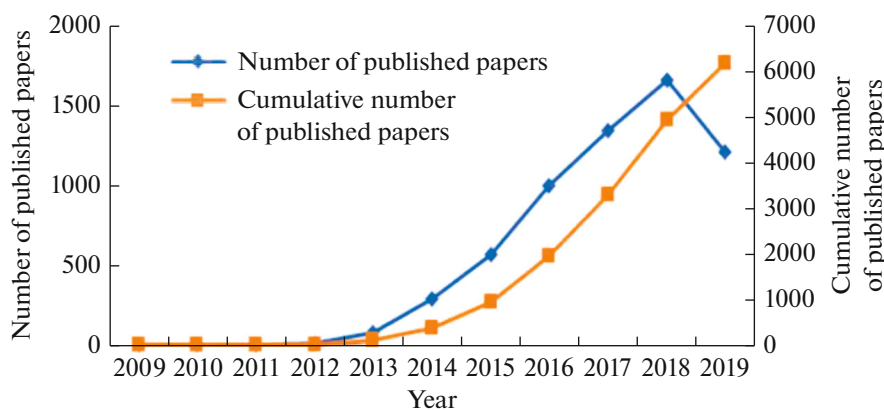


Fig. 1. Changes in the number of annual publications on the research of CRISPR-editing indexed in the Web of Science during 2009–2019.

each year from 2009 to 2011. The number of articles showed a steady upward trend in the following years, and there was a sharp increase in 2018. The reason for the lower number of releases in 2019 compared with 2018 is that data are not annual.

Analysis of cited references with high citation frequency. Through the co-citation analysis of literature, alternative scientific literature is selected to provide other researchers with more authoritative knowledge. According to WoS results, the top 10 articles with the highest citation frequency were obtained (Table 1). There are Cong L, Jinek M, Mali P, Ran FA, Hsu PD, Doudna JA, Wang HY, Gaj T, and Qi LS. Cong L ranks first in the cited frequency. Hsu PD has two articles, including DNA targeting specificity of RNA-guided Cas9 nucleases published in 2013 and Development and Applications of CRISPR-Cas9 for Genome Engineering published in 2014.

Co-citation analysis. Co-citation analysis of literature is the most prominent feature of CiteSpace. Mining literature co-citation relationships through a literature spatial data set then show the network distribution. The condition for co-citation of two papers is that they appear in the reference list of the third cited paper. After analysis, the number of nodes is 256, and the number of lines is 402 (Fig. 2). This indicates that there are 402 pairs of co-cited relationships in 256 articles, and the research content is similar. Cong L has the largest node and the highest citation, but Jinek M ranks first in centrality. This indicates that these articles have significance and representative of CRISPR gene-editing.

Knowledge Subject Analysis of CRISPR Gene-editing

Correlation analysis of core authors. Continuous development is closely related to the efforts of researchers. According to the statistics, a total of 251 authors participated in the CRISPR gene-editing

study (Fig. 3). It is shown that person with the largest purple circle is Jinek M (2449), followed by Cong L (2410), Mali P (2231), Ran Fa (1865), and Hsu Pd (1759), which reflected their strongest academic authority. However, the authority should be judged in combination with centrality, not just the number of papers published. Co-citation refers to two (or more authors) being cited by one or more subsequent papers at the same time, which proves that these two or more authors constitute co-citation relationship. The nodes represent the authors, there are many collaborative clustering systems with the close connection between nodes indicates that close cooperation. The author co-occurrence network can indicate which authors are highly influential. The authors with close connections have similar research topics, and the contacts with different topics have multidisciplinary backgrounds or interdisciplinary research fields. The density value is 0.0123, The higher the density value is, the better the clustering result of the network is. There is no exact standard (Fig. 4). There are relatively large cooperative groups mainly including Zhang Y, Zhang F, Wang Y, Kim JS, and Liu Y in the network. Small groups surrounding large groups indicate that the research system is relatively complete, and less influential ones will focus on more influential ones for further research.

Correlation analysis of core institutions. All literature retrieved involved 147 institutions, three institutions have published more than 100 papers (Table 2). The United States has seven and China has three among the top ten institutions. Universities are the mainstay of research institutions, accounting for 80%. The largest number is the Chinese Academy of Sciences (233). Subsequently, Harvard Univ (103), Univ Calif Berkeley (102) ranked 2nd, 3rd. Stanford Univ's centrality (0.39) ranks first, which indicates that it has made outstanding contributions in the field of genetic engineering.

Table 1. The top 10 literature statistics

	Literature	Cited
1	Cong L, Ran FA, Cox D, Lin SL, Barretto R, et al. Multiplex Genome Engineering Using CRISPR/Cas Systems SCIENCE. 2013 FEB 15; 339 (6121): 819–823	5452
2	Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, et al. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity SCIENCE. 2012 AUG 17; 337 (6096): 816–821	4302
3	Mali P, Yang LH, Esvelt KM, Aach J, Guell M, et al. RNA-Guided Human Genome Engineering via Cas9 SCIENCE. 2013 FEB 15; 339 (6121): 823–826	3886
4	Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, et al. Genome engineering using the CRISPR-Cas9 system NATURE PROTOCOLS. 2013 NOV; 8 (11): 2281–2308	2781
5	Hsu PD, Lander ES, Zhang F Development and Applications of CRISPR-Cas9 for Genome Engineering CELL. 2014 JUN 5; 157 (6): 1262–1278	1939
6	Doudna JA, Charpentier E The new frontier of genome engineering with CRISPR-Cas9 SCIENCE. 2014 NOV 28; 346 (6213): Art. No. 1258096	1770
7	Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, et al. DNA targeting specificity of RNA-guided Cas9 nucleases NATURE BIOTECHNOLOGY. 2013 SEP; 31 (9): 827–+	1716
8	Wang HY, Yang H, Shivalila CS, Dawlaty MM, Cheng AW, et al. One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering CELL. 2013 MAY 9; 153 (4): 910–918	1675
9	Gaj T, Gersbach CA, Barbas CF ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering TRENDS IN BIOTECHNOLOGY. 2013 JUL; 31 (7): 397–405	1488
10	Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, et al. Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression CELL. 2013 FEB 28; 152 (5): 1173–1183	1479

The atlas contains 147 nodes and 105 lines (Fig. 5). There is less connection between the 147 agencies and the cooperation is regional. There is a slightly more internal connection between institutions in the domestic.

Correlation analysis of core countries. The analysis of core countries can reflect the overall structure of a certain research field and clarify the influence of different countries and the cooperative relationship between countries. The contribution of countries was observed (Table 3). The USA led with 2631, followed by China (1304), Japan (411) and Germany (362), and England (284). The number of publications is inconsistent with the centrality. Analysis of cooperation between countries is shown in Fig. 6. The atlas ($N = 62$, $E = 56$, Density = 0.0296) show that 62 countries and

56 cooperative relationships have been formed. USA is investing more effort and is cooperating with many countries. However, China has connections with only a few countries in terms of connectivity. These results show that although China has published a large number of papers in the field of CRISPR gene-editing, there is still a lack of cooperation with other countries.

Correlation analysis of core journals. Researchers have published 6150 articles in 1190 journals. The ranking is based on the number of articles published. The results of the top 10 journals are shown in Table 4, the highest cited journal is SCIENCE, followed by NATURE, P NATL ACAD SCI USA, and NAT BIOTECHNOL. Researchers were able to publish massive in such a short period in authoritative journals, which

Copyright © 2021
 BIOLOGY BULLETIN
 ISSN 1841-2716
 DOI: 10.2478/BB-2021-0001
 Published online
 2021, Volume 48, No. 6
 www.biologybulletin.com

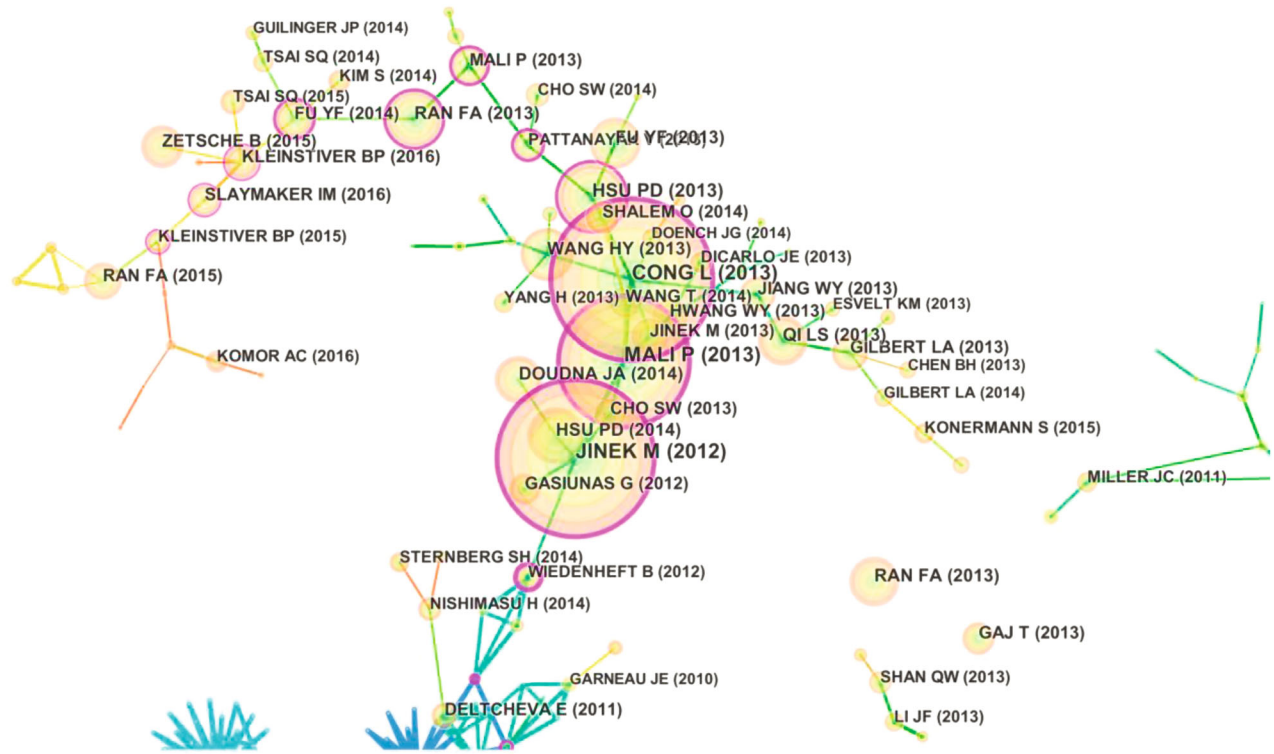


Fig. 2. Visualization based on co-citation network: The nodes represent the authors. The radius of circles is in proportion to the amount of literature of the category, the larger the node, the more frequently it is referenced. Important nodes are highlighted with purple rings.

is enough to illustrate the potential, usefulness, and impact of CRISPR gene-editing.

Hot spot analysis based on high-frequency words.

Keywords are extracted from the title and abstract. Keywords highly summarize the content and reflect the main content and core theme of the literature. A

visual knowledge map of keywords with high frequency could reflect hot topics. Among the node labels, “Cas9”, “CRISPR/Cas9”, “CRISPR”, “system”, and “expression” (Table 5) ranked ahead in both the frequency list, which suggested that they were the hotspots, as well as an important turning point.

Table 2. The table of organization that published the articles

Freq	Centrality	Year	Institution
233	0.28	2013	Chinese Acad Sci
103	0.16	2013	Harvard Univ
102	0.03	2012	Univ Calif Berkeley
93	0.15	2013	MIT
87	0.39	2014	Stanford Univ
73	0	2014	Harvard Med Sch
69	0	2014	Univ Chinese Acad Sci
60	0.14	2013	Duke Univ
59	0.23	2013	Univ Calif San Francisco
59	0.08	2016	Chinese Acad Agr Sci

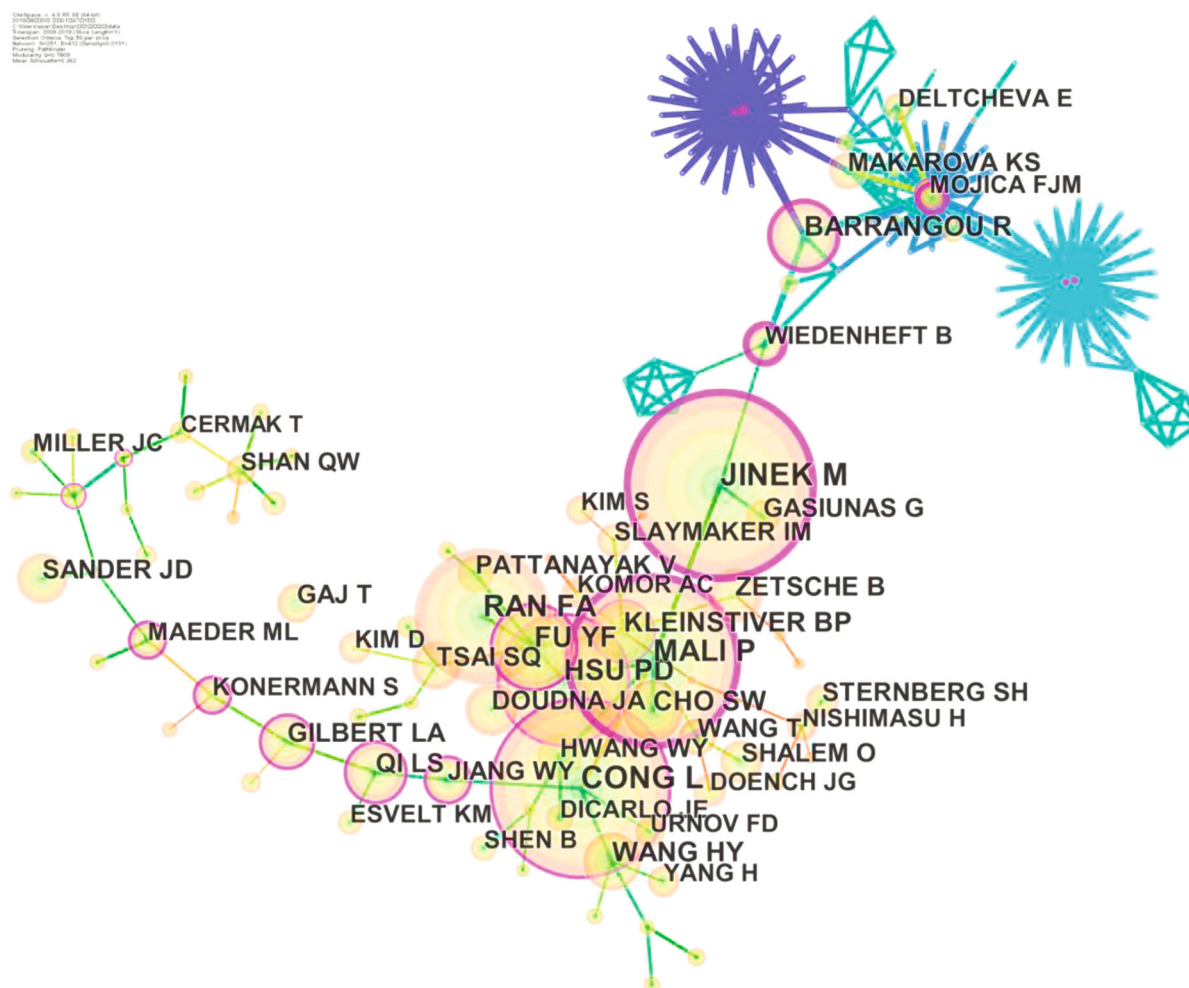


Fig. 3. Graph of co-cited authors: A node represents an author, and the node size represents frequency.

Keywords are further clustered, and keyword co-occurrence clustering analysis can identify the relationship and find hot topics. Figure 7 is obtained by the TF-IDF method (Kardkovács and Kovács, 2015) which can evaluate the importance of research objects

to the whole field. Clustering module value (Q value) > 0.3 means that the clustering structure is significant. Clustering average contour value (S value) > 0.5 means that clustering is reasonable. $Q = 0.8463$ and $S = 0.594$ are shown in the figure, results are highly clustered and

Table 3. Information table issued by Countries

Freq	Centrality	Year	Country/Region
2631	0.63	2009	USA
1304	0	2013	PEOPLES R CHINA
411	0	2013	JAPAN
362	0.12	2013	GERMANY
284	1.08	2013	ENGLAND
219	0.06	2013	SOUTH KOREA
156	0.6	2012	CANADA
146	0.12	2012	FRANCE
131	0.06	2013	NETHERLANDS
119	0.06	2013	INDIA

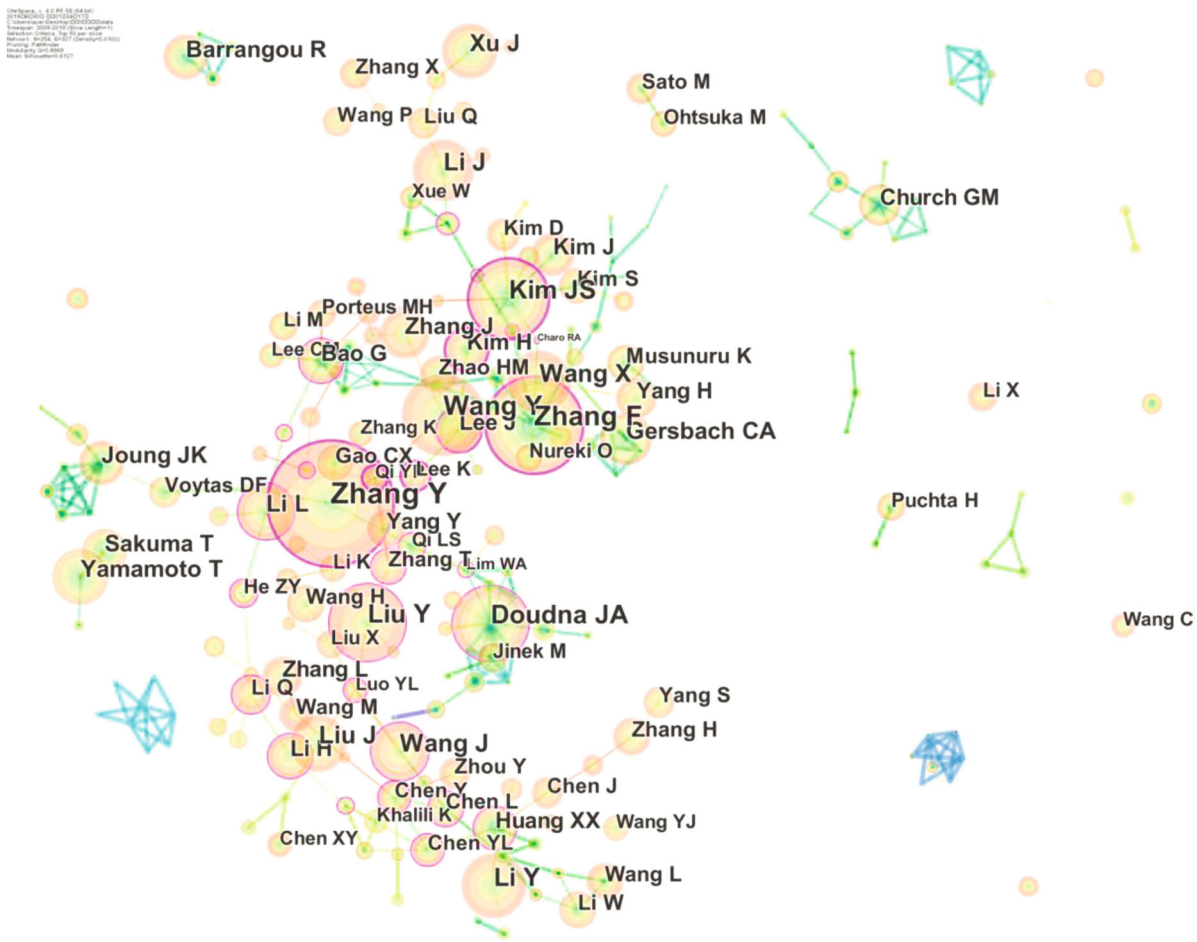


Fig. 4. The cooperation network of productive authors: a node represents an author, and a line represents a connection.

convincing. Different clusters with different colors and shapes form eight keyword co-occurrence network clustering. Clusterings #0-#7: Endoribonuclease, *Caenorhabditis Elegans*, Cpf1, CRISPR/cas9 system, identification, *saccharomyces cerevisiae*,

community, and so on, a cluster is a research field, and literature within a cluster are closely linked. Endoribonuclease (#0) is the largest cluster with the highest citation frequency. Nodes in one cluster overlap with nodes in another cluster, indicating that the content is

Table 4. Table of cited journals

Freq	Centrality	Year	Journal
5166	0.52	2009	SCIENCE
4794	0.11	2009	NATURE
4604	0.35	2009	P NATL ACAD SCI USA
4546	0.07	2011	NAT BIOTECHNOL
4449	0.04	2009	CELL
3863	0	2009	NUCLEIC ACIDS RES
3409	0	2013	PLOS ONE
3167	0	2013	NAT METHODS
2747	0	2014	SCI REP-UK
2558	0	2013	NAT COMMUN



Fig. 5. Institutional Collaboration Graph: A node represent an institution, and node size represents frequency.

interrelated. The same keywords appear in different clusters, indicating that they are being followed in different periods. When a group of keywords appears in the same group of literature at the same time, which shows that they have a co-occurrence relationship.

The more co-occurrence frequency is, the more similar the research topic is.

The frontier of CRISPR gene-editing research. “Mutation detection” can identify sudden changes in the information, the burst strength of mutant words

Table 5. Keywords information table

Freq	Centrality	Keyword	Year	ClusterID
1109	0.31	cas9	2013	3
1085	0	crispr/cas9	2013	7
1058	0.07	crispr	2013	3
861	0.89	system	2012	0
858	0	expression	2013	3
832	0.12	genome editing	2013	3
793	0	gene	2013	0
761	0	crispr cas9	2015	3
696	0	genome	2011	7
690	0.44	dna	2010	4

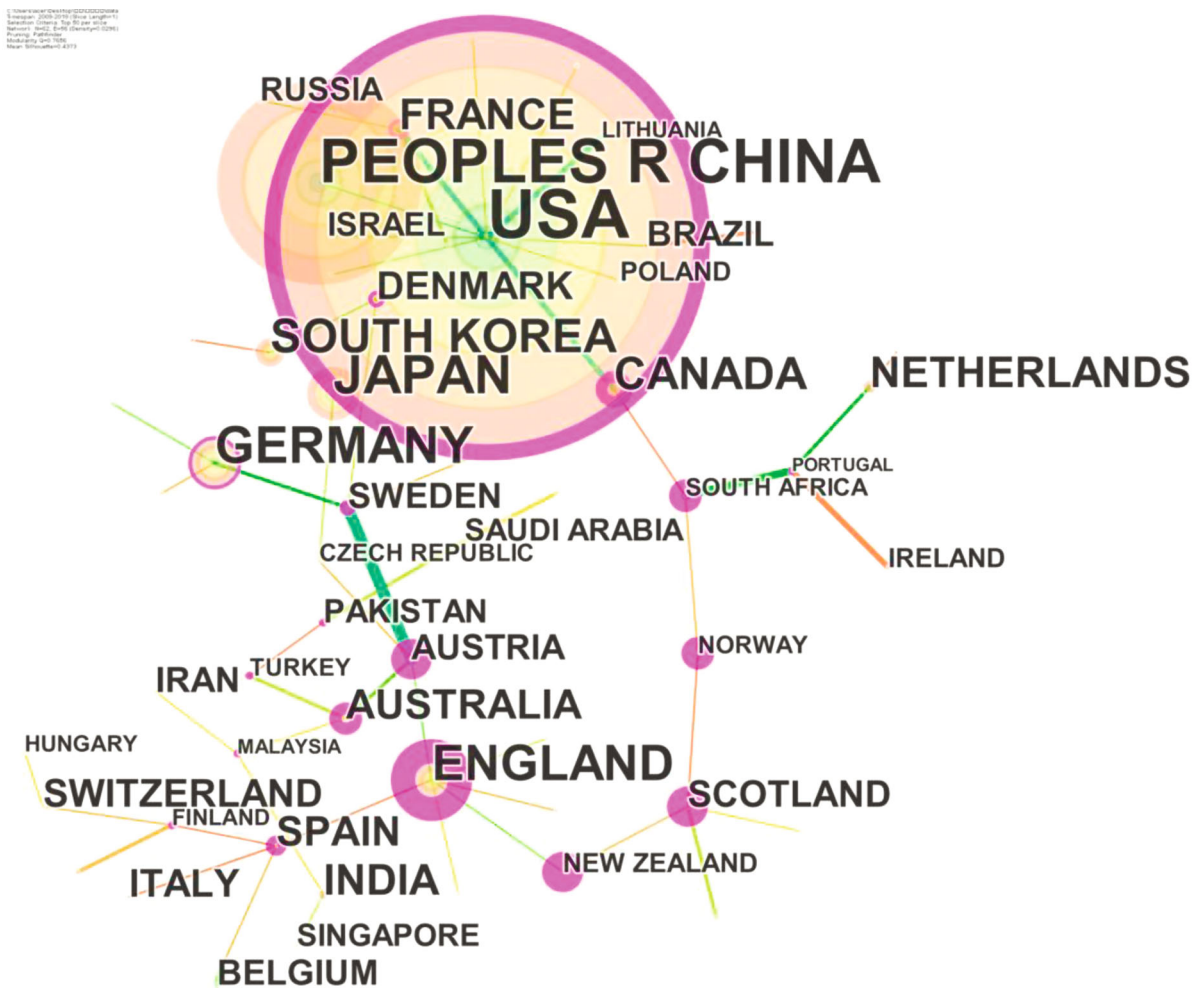


Fig. 6. Map of country cooperation: A node represents a country, the lines reflect the relationships of collaboration, more lines, higher the times of co-citation. Thicker lines, closer the partnership.

can reflect the transfer of research hotspots and frontiers. Table 6 sorted the 138 mutant words according to the intensity of mutation and listed 40 with mutation intensity > 1. After eliminating the ultra-low frequency vocabulary such as transferred gene, stationary phase and genome-wide hypermutation, the top ten are followed by “zinc finger nucleases” (2013–2016), “one-step generation” (2014–2015), “specificity” (2014–2015), “talen” (2013–2014), “bacteria” (2010–2014), “human cell” (2014–2015), “*C. elegans*” (2013–2014), “*Drosophila*” (2013–2015), “interference” (2012–2014), and “CRISPR/Cas9” (2016–2017).

Keywords time zone maps will be cited in the form of timelines and reveal the historical trajectory, which can be used to analyze research hotspots and emerging research directions in different periods. According to Fig. 8, the keywords that appeared in 2015 are “mammalian cell”, etc. The keywords that appeared in 2016 are “CRISPR/Cas9 system”, etc. The keywords that appeared in 2017 are “*in vitro*”, etc. The keywords that appeared in 2018 are “*Arabidopsis*, rice, synthetic biology, disease”, etc. The keywords that appeared in 2019

are “*cpf1*, model, therapy”, etc. CRISPR has developed rapidly in recent years.

DISCUSSION

Growth Trend of Annual Publication

By this result shows that in recent years the CRISPR gene-editing study has the following characteristics: we analyzed the published literature, the literature data in 2019 were excluded, the number has been at a high level and rising trend year by year. In the foundation stage (2009–2011), the CRISPR system is starting to sprout. As the research deepens, more findings emerged in the later period. During the flourishing period stage (2012–2019), there was a sharp growth of CRISPR-related publications. In 2012, Doudna JA (Jinek et al., 2012) demonstrated that CRISPR can be used as a tool for gene-editing, CRISPR has rapidly become the most popular gene-editing tool in human biology, agriculture, microbiology, and other fields.

bioRxiv preprint doi: <https://doi.org/10.1101/20210304>; this version posted March 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

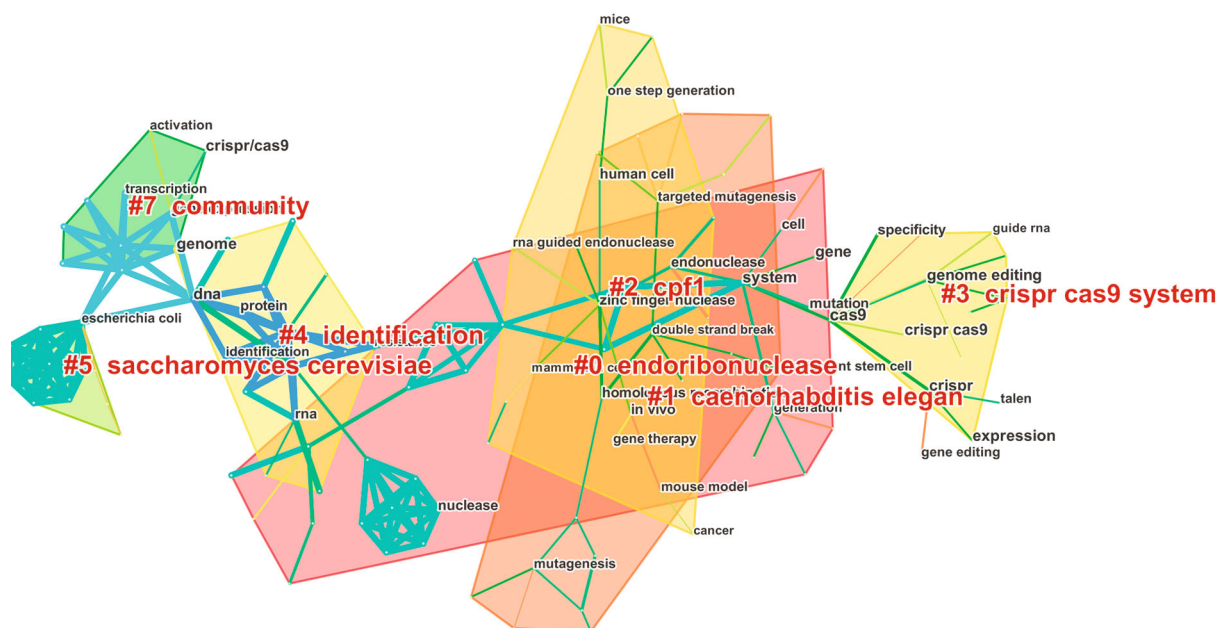


Fig. 7. Keywords cluster map (TF-IDF): A shape represents a cluster, color means the time of the first occurrence, the color gradually changes from cool to warm, blue indicates the earlier year, red indicates the warmest color for the most recent year. The smaller the cluster tag number, the higher the number of published literature and the frequency of citation.

F

bioRxiv preprint doi: <https://doi.org/10.1101/20210304>; this version posted March 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

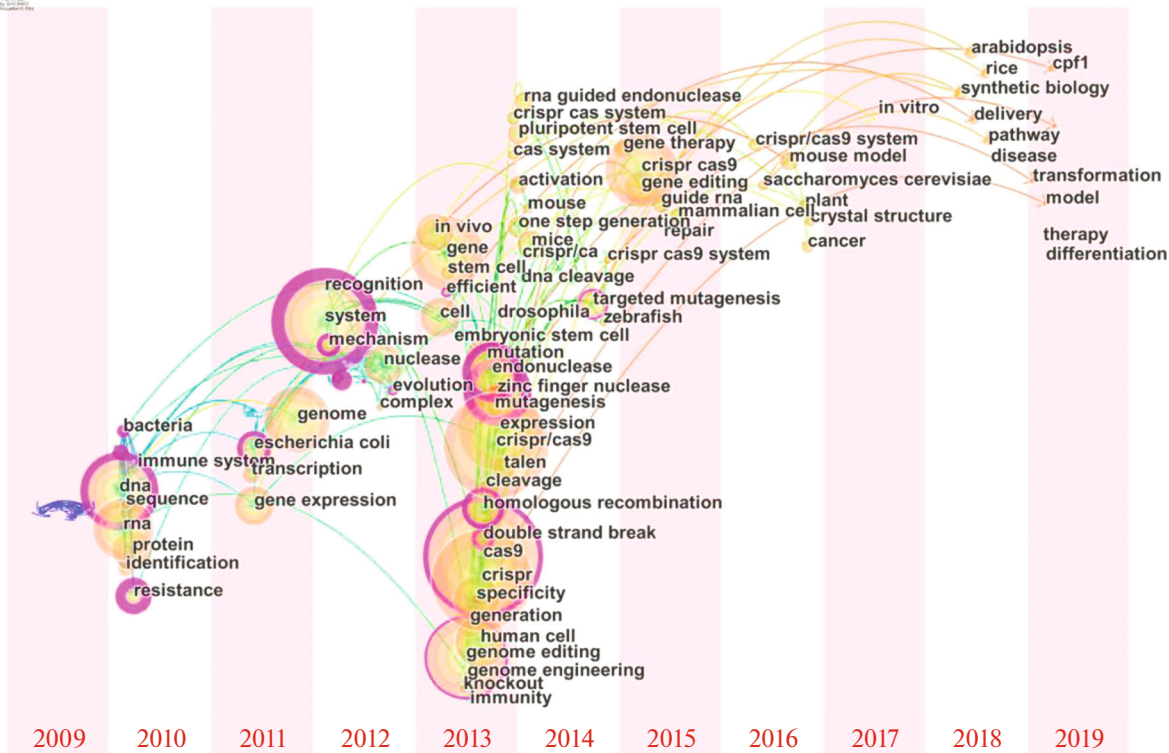


Fig. 8. Keywords time zone diagram (timeline view): The heavier color is displayed, and closer the time is. The lines can also be seen that there is a clear relationship between the studies of different stages, indicating research is progressive.

Table 6. Top 40 Keywords with Strongest Citation Bursts

Keywords	Year	Strength	Begin	End	2009–2019
Zinc finger nuclease	2009	44.6836	2013	2016	
One step generation	2009	26.7624	2014	2015	
Specificity	2009	21.1505	2014	2015	
Talen	2009	15.0461	2013	2014	
Bacteria	2009	14.0849	2010	2014	
Human cell	2009	13.4507	2014	2015	
Caenorhabditis elegans	2009	11.9377	2013	2014	
Drosophila	2009	9.3779	2013	2015	
Interference	2009	9.1783	2012	2014	
Crispr/cas9 system	2009	9.1568	2016	2017	
Immune system	2009	9.1096	2010	2013	
Crispr/ca	2009	8.5779	2014	2015	
Tal effector	2009	7.6477	2013	2014	
Mammalian cell	2009	7.3572	2015	2016	
Cas system	2009	7.2893	2014	2015	
Endonuclease	2009	7.0987	2013	2014	
C. elegans	2009	7.0556	2013	2015	
Human genome	2009	6.7194	2013	2014	
Embryo microinjection	2009	6.5814	2013	2015	
Antiviral defense	2009	6.4916	2012	2014	
Archaea	2009	6.0929	2013	2014	
Knockout rat	2009	6.0162	2013	2014	
Gene targeting	2009	5.6279	2013	2014	
Adaptive immunity	2009	5.6069	2013	2015	
Crispr cas system	2009	5.494	2015	2016	
Defense	2009	5.468	2010	2013	
Sequence	2009	4.7997	2010	2013	
Prokaryote	2009	4.404	2012	2015	
Homologous recombination	2009	4.2677	2013	2015	
Crispr rna	2009	4.1661	2012	2014	
Multiplex	2009	3.7566	2013	2014	
Complex	2009	3.5699	2012	2013	
Streptococcus thermophilus	2009	3.1506	2012	2014	
Bacteriophage	2009	3.0339	2012	2013	
Seed sequence	2009	2.4453	2012	2014	
Resistance	2009	2.1789	2010	2012	
Recognition	2009	1.9026	2012	2013	
Short palindromic repeat	2009	1.8704	2012	2014	
Provides acquired resistance	2009	1.6019	2009	2014	
Non coding rna	2009	1.2756	2012	2013	

Analysis of High-impact Publication

In the high literature co-citation analysis, the literature with high frequency and centrality reveals the hot spots. For example, the Multiplex-genome Engineering citation of the CRISPR / Cas system pub-

lished by Zhang F (Cong et al., 2013) in the 2013 citation outbreak, this article is the first to achieve gene-editing in mammalian cells and implemented several sites simultaneously for editing. The frequency of citations published by the Doudna JA ranked second, but

they first revealed that the CRISPR system can be used for genome editing, this article has set off a huge wave.

Analysis of Co-Cited Authors

In knowledge subject analysis, the collaboration patterns among the core authors with the strongest academic authority were further analyzed. The total citation frequency is not proportional to the centrality but most of the authors have both high citation frequency and high centrality. Zhang F not only initiated the application of CRISPR/Cas9, but added CRISPR/Cas12a, CRISPR/Cas13, and so on. In 2020, Jennifer Doudna and Emmanuelle Charpentier won the Nobel Prize in chemistry for inventing CRISPR-Cas9 and made immeasurable contributions to mankind. The mapping shows the emergence of many researchers, but few productive authors. Collaborations among authors should be encouraged to accelerate the pace and depth of research in this field.

Analysis of Leading Countries and Institutions

In terms of co-occurrence frequency and centrality, the countries with a high impact are mainly developed countries in North America and Europe, including the USA, Germany, England, Canada, France, and the Netherlands. Asia mainly including China, Japan, South Korea, and India. Regarding institutional analysis, the country with the largest number of institutions is still the USA. Chinese are relatively late to CRISPR, but second only to the USA in terms of publication. However, the centrality of Chinese studies is close to zero. The USA has a wide network of partnerships. Although internal institutions are closely connected, the overall form of cooperation between countries and institutions is relatively scattered.

Analysis of Co-Cited Journals

In the analysis of co-cited journals, SCIENCE and NATURE ranked first. This study found that the current top journals remain consistent, but the number of articles published in journals focusing on genes (including DNA, RNA) and gene-editing technology has relatively increased, especially NAT BIOTECHNOL and NUCLEIC ACIDS RES are included. The possible reason is that the number of publications is exploding and professional journals are increasingly promoting sustainability. The fact that CRISPR has been published in top international journals, including Science and Nature, is a testament to its significance.

The Hotspots and Frontiers of CRISPR

For the detection of research frontier and hotspots, Citespace IV detects new trends and sudden changes in the development of the subject by extracting emer-

gent subject terms, and solves the defect of relying solely on the statistic of the word frequency of subject words well through dynamic network aggregation. Shifting frontiers and hot spots can clearly indicate which way the CRISPR editing “winds” are blowing, providing researchers with some new ideas. On the analysis of research frontiers, the development process of this discipline can be divided into three parts. The first stage that known as the exploratory stage was from 2009 to 2011, Mutants mainly included bacteria, immune system, Escherichia coli, etc. Indicating that researchers were mainly exploring the specific mechanism of CRISPR in this period. The second phase was from 2013 to 2015 when there was a significant increase in the number of published papers. Keywords such as Crispr /cas9, human cells, gene editing, and mice started to burst in 2013 which represented scientists have applied CRISPR as a gene-editing tool to all kinds of organisms. The third phase was from 2016 to 2019, the emergence of disease, cancer, and therapeutic differentiation during the advanced phase suggests that researchers are working to apply the technology to clinical treatment and even attack cancer.

Application of CRISPR in plants, animals. In hot and frontier analysis, the frequency and centrality of cas9 ranked among the top, indicating that CRISPR/Cas9 which is simpler in structure, easier to operate, and more feasible is a great focal point. According to the analysis, the CRISPR/Cas9 study mentioned bacteria such as escherichia coli slightly more frequently, and then zebrafish, drosophila, plant, c. elegans, human cell, in vivo, and other words began to appear. Indicating that CRISPR application research is the mainstream direction, the focus of research has been transferred to plants and animals, such as zebrafish (Sharma et al., 2021), pig (Tanihara et al., 2021), tobacco (Zhang et al., 2021), arabidopsis, wheat, and rice.

Discovery and application of other representative nucleases. After the keyword clustering analysis, new words such as Cpf1 and other words represented the trends. With the development of CRISPR gene-editing, scientists have found a more convenient system-CRISPR/Cpf1 (type V CRISPR/Cas12a). Zhang F (Makarova et al., 2020) pointed out that Cpf1 is simpler than the Cas9 system. Cpf1 requires only one RNA to cut the target gene. The Cpf1 enzyme is also smaller than standard spCas9, making it easier to deliver to cells and tissues. The Cpf1 complex leaves an overhang at the exposed end, which is expected to make the integration more accurate. The incision of Cpf1 is away from the recognition site, providing multiple opportunities for correction editing. Then in 2018, B HL (Harrington et al., 2018) discovered Cas14-the smallest functional CRISPR system which is a family of exceptionally compact RNA-guided nucleases (400 to 700 amino acids). Cas14 proteins are

capable of targeted single-stranded DNA (ssDNA) cleavage without a restrictive sequence. Moreover, target recognition by Cas14 triggers nonspecific cutting of ssDNA molecules, an activity that enables high-fidelity single-nucleotide polymorphism genotyping (Cas14-DETECTOR). In 2019, Jun (Liu et al., 2019) has isolated one of these nucleases and named CasX which is much smaller and has features not found in other Cas proteins such as a domain involved in DNA unwinding. It has a unique programmable editing approach that may have advantages that the current CRISPR Cas genome editing technology does not. The most commonly used tools are SpCas9 and Cas12a (Cpf1) at present.

Application of CRISPR in clinical treatment. CRISPR is used not only for genetic modification and gene therapy but for other diseases. China is one of the first countries to use CRISPR in human trials. Limsirichai (Limsirichai et al., 2016) used the acetyltransferase-based CRISPR system to mediate the upregulation of HIV-1 integrated gene expression. H. C. Yang (Yang et al., 2014) co-transfected HBV-expressed plasmid and gRNA-Cas9 plasmid into Huh-7 cells, and found that the expression of the HBV gene was significantly suppressed. Besides, Tiffany (Russell et al., 2015) and Xu (Xu et al., 2015) also respectively made corresponding studies on herpes simplex virus and rabies virus. Advances have even been made in the treatment of parasites such as *Toxoplasma* (Shen et al., 2014) and *Trypanosoma* (Lander et al., 2015). At the same time, research has shown that there is high editing efficiency in *S. pneumoniae* and *E. coli*, and even mutation efficiency may be reached 100% in *S. pneumoniae*. This technology can also be used to improve the ability of fermentation strains to resist phage infection in the process of food fermentation production. Besides, breakthroughs have also been made in the treatment of other intractable diseases such as cataract genetic (Wu et al., 2013) and neurological (Mohamed et al., 2019) diseases, and even cancer.

CRISPR/Cas9 technology has played an important role in clinical tumor treatment including CAR-T cell immunotherapy, immune checkpoint blocking therapy, and antibody-targeted therapy. The use of CRISPR/Cas9 technology in antibody-targeted therapies, targeted editing of hybridoma cells, and B cell immunoglobulin heavy chain constant region gene can achieve type transformation of antibodies to obtain specific classes (Anguille et al., 2015). In 2015, Aubrey (Aubrey et al., 2015) applied CRISPR/Cas9 gene-editing technology to malignant lymphoma for the first time. Subsequent series of validations proved that lentivirus-mediated CRISPR/Cas9 can be used as a novel tumor treatment scheme. Zhang F (Maeder et al., 2019) treated LCA10 by injecting the AAV5 vector into the subretinal of mice and using double gRNA to target the upstream and downstream, achieved the

overall deletion or inversion of the mutant region. The study has completed the first phase I/II clinical administration of patients in 2020 and is the world's first in vivo administration of CRISPR gene-editing. Researchers have also made breakthroughs in cancer research such as cervical cancer, ovarian cancer, osteosarcoma, prostate cancer (Wei et al., 2018), melanoma (Nagler et al., 2020), leukemia (Narimani et al., 2019), and pancreatic cancer (Zhao et al., 2018). However, most studies can only inhibit but not destroy tumors, the off-target rate is lower in normal embryonic stem cells and normal animal model embryos and higher in cancer cell systems.

Teams from the Karolinska Institute in Sweden and the Novartis Institute for Biomedicine in Cambridge found that CRISPR-edited cells often have a defect in the tumor suppressor gene p53, and clinical treatment with CRISPR/Cas9 may increase the risk of cancer (Haapaniemi et al., 2018). Another team identified a wide range of genotypes with the third-generation sequencing technology and found that CRISPR/Cas9 may cause the loss of large genomic fragments near the target, as well as problems of DNA rearrangement but no corresponding report has been reported at the embryo level. In summary, CRISPR is not a mature genetic tool, particularly off-target effects, which have been addressed.

CRISPR detection platform. In vitro DNA detection system, A pair of dCas9 proteins are connected to half-split fluorescein and dCas9 and a single microring resonator biosensor combination technology enables have enabled the real-time, label-free detection of pathogenic DNA and RNA. Besides, a biosensor called CRISPR-Chip utilizes the targeting ability of dCas9. It is a device that combines with specific sgRNA and is fixed on a transistor to generate a label-free nucleic acid detection device with high sensitivity and fast detection speed. After David (Nelles et al., 2016) added a PAMmer oligonucleotide sequence, dCas9 can bind RNA to the subcellular localization of RNA in living cells. In recent years, a technology called "CASFISH" has been developed (Deng et al., 2015), which combines dCas9 with FISH (fluorescence in situ hybridization) to avoid the natural structural deformation of the nucleus caused by heating and destroying the chemicals in FISH. The CASFISH analysis of the dCas9/sgRNA complex can produce multi-color markers that can be used for target loci in cells. In addition, the CASFISH test is very fast under optimal conditions and can be used to detect major tissue sections.

Cas13 specifically cuts RNA rather than DNA, so it usually has important application value in molecular diagnosis and down-regulation of gene expression. Zhang F (Joung et al., 2020) reported a detection technology based on Cas13-SHERLOCK. Later, they

described a simple test for the detection of SARS-CoV-2. STOP (SHERLOCK testing in one pot) that combines simplified extraction of viral RNA with isothermal amplification and CRISPR-mediated detection is a streamlined assay that (Joung et al., 2020). Parson Sabeti and Zhang F (Freije et al., 2019) built a powerful and fast programmable diagnostic and antiviral system-CARVER according to the antiviral activity and diagnostic capabilities of Cas13, which can target multiple ssRNA viruses to detect and eliminate human cells RNA virus. The system programmatically cuts RNA that is complementary to crRNA and may be used to diagnose and treat viral infections in the future. If further combined with the dCas9 system to achieve RNA Simultaneous labeling of transcription and gene loci provides a way to study RNA positions, the interrelationships between RNAs, and the relationship between DNA and RNA transcription regulation in living cells. The DETECTOR diagnostic tool binds and cleaves the amplified product through Cas12a / crRNA to activate DNase activity, which releases a fluorescent signal after the single-stranded DNA-conjugated fluorescent reporter is cleaved. High-risk Types 16 and 18 HPV have been detected (Li and Li, 2020). When Cas12a and Cas13 systems both are added at the same time, it is found that the Cas enzyme corresponding to the target gene will turn on the cleavage activity and release the corresponding fluorescent reporter group to detect the ssRNA target and dsDNA target simultaneously in a single reaction.

Strengths and Limitations

The research field of CRISPR could be regarded as an attractive and emerging field. Although over the past few years, we've gone through three periods of exploration, prosperity, and depth, learned more about CRISPR/Cas9 gene-editing technology, Clinical treatment is still an intractable difficulty for humans. Thus, it makes sense to carry out a visualization study on CRISPR-related research. This study expressed a series of articles, complex subjects, or knowledge through graphs and tables with the map method of computer information visualization, and the development history, cooperative relations, and research frontiers of subjects can be visually and vividly displayed. This study is comprehensive, the analysis of core journals and research hotspots was added. The number of published papers in the field of CRISPR has shown a 'blowout' trend. Additionally, we could observe the crucial landmark papers with brief article information. Compared with other methods such as Bibexcel (Farzanegan et al., 2017), although the Bibexcel selects a database that can well represent scientific articles, it is poor in analyzing some research results (such as patents). Citespace even

realizes a separate patent metering, which has a wider range of applications.

As this study is based on the literature, several major limitations are also considered. Firstly, some specified introduction but not the full text due to its analysis feature, visualization figures could lack some crucial details when constructed. Besides, we only collected data from Web of Science and set the length of time as the most recent 10 years when exploring literature, which might lead to bias.

CONCLUSIONS

CiteSpace can be seen as a powerful tool in the era of big data. People can directly see the analysis results through figures and tables through visual analysis of data. However, there are some shortcomings, due to the use of purely bibliometric data analysis in this method, the medical analysis results obtained are limited, which may lead to insufficient comprehensive analysis results and other problems.

This study shows that CRISPR is a milestone for human genetic engineering. In recent years, the research on CRISPR gene-editing in China has developed rapidly, but the quality still needs to be improved. It is suggested that close cooperation groups should be established with research institutions in other countries to establish perfect research systems and jointly promote the development of genetic engineering.

ACKNOWLEDGMENTS

We acknowledge the authors.

FUNDING

This work was partly supported by The National Nature Science Foundation of China (Grant No. 81860653,82060654); The Science and Technology Foundation of Shaanxi Province [2020JM-550, 2020JM-545 and 2021JM-416]; Open Project Foundation of Key Laboratory of noncoding RNA and drugs in Universities of Sichuan Province [FB19-01 and FB20-02]; Shaoxing Medical Key Discipline Construction Plan (NO. 2019SZD06); Project of Clinical Research Fund Project of Zhejiang Medical Association (2020ZYC-A125) Yangtze River Research Project for Sustainable Development of Hospitals in Zhejiang Province (2020ZHA-YZJ218).

COMPLIANCE WITH ETHICAL STANDARDS

The authors declare that they have no conflicts of interest. This article does not contain any studies involving animals or human participants performed by any of the authors.

REFERENCES

- Anguille, S., Smits, E.L., C. Bryant, H. H. Van Acker, H. Goossens, E. Lion, P. D. Fromm, D. N. Hart, V. F. Van Tendeloo, and Z. N. Berneman. 2015. Dendritic Cells as Pharmacological Tools for Cancer Immunotherapy. *Pharmacol Rev* 67:731–753.
- Aubrey, B. J., Kelly, G.L., Kueh, A.J., Brennan, M.S., O'Connor, L., Milla, L., Wilcox, S., Tai, L., Strasser, A., and Herold. M.J., An inducible lentiviral guide RNA platform enables the identification of tumor-essential genes and tumor-promoting mutations in vivo, *Cell Rep.*, 2015, vol. 10, pp. 1422–1432.
- Carter, J. and Wiedenheft. B., SnapShot: CRISPR-RNA-guided adaptive immune systems, *Cell*, 2015, vol. 163.
- Chen, C., CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature, *J. Am. Soc. Inform. Sci. Technol.*, 2006, vol. 57.
- Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., and Zhang, F., Multiplex genome engineering using CRISPR/Cas systems, *Science*, 2013, vol. 339, pp. 819–823.
- Deng, W., Shi, X., Tjian, R., Lionnet, T., and Singer. R.H., CASFISH: CRISPR/Cas9-mediated in situ labeling of genomic loci in fixed cells, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, vol. 112, pp. 11870–11875.
- Farzanegan, R., Feizabadi, M., Ghorbani, F., Movassaghi, M., Vaziri, E., Zangi, M., Lajevardi, S., and Shadmehr. M.B., An overview of tracheal stenosis research trends and hot topics, *Arch. Iran. Med.*, 2017, vol. 20, pp. 598–607.
- Freije, C.A., Myhrvold, C., Boehm, C.K., Lin, A.E., Welch, N.L., Carter, A., Metsky, H.C., Luo, C.Y., Abudayyeh, O.O., Gootenberg, J.S., Yozwiak, N.L., Zhang, F., and Sabeti. P.C., Programmable inhibition and detection of RNA viruses using Cas13, *Mol. Cell*, 2019, vol. 76.
- Haapaniemi, E., Botla, S., Persson, J., Schmierer, B., and Taipale. J., CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response, *Nat. Med.*, 2018, vol. 24, pp. 927–930.
- Harrington, L.B., Burstein, D., Chen, J.S., Paez-Espino, D., Ma, E., Witte, I.P., Cofsky, J.C., Kyrpides, N.C., Banfield, J.F., and Doudna. J.A., Programmed DNA destruction by miniature CRISPR-Cas14 enzymes, *Science*, 2018, vol. 362, pp. 839–842.
- Heler, R., Marraffini, L.A., and Bikard. D., Adapting to new threats: the generation of memory by CRISPR-Cas immune systems, *Mol. Microbiol.*, 2014, vol. 93, pp. 1–9.
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier. E., A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity, *Science*, 2012, vol. 337, pp. 816–821.
- Joung, J., Ladha, A., Saito, M., Kim, N.-G., Woolley, A.E., Segel, M., Barretto, R.P.J., Ranu, A., Macrae, R.K., Faure, G., Ioannidi, E.I., Krajewski, R.N., Bruneau, R., Huang, M.-L.W., Yu, X.G., Li, J.Z., Walker, B.D., Hung, D.T., Greninger, A.L., Jerome, K.R., Gootenberg, J.S., Abudayyeh, O.O., and Zhang, F., Detection of SARS-CoV-2 with SHERLOCK one-pot testing, *N. Engl. J. Med.*, 2020, vol. 383, pp. 1492–1494.
- Kardkovács, Z.T., and Kovács. G., *Finding sequential patterns with TF-IDF metrics in health-care databases, Acta Universitatis Sapientiae, Informatica*, 2015, vol. 6.
- Keeney, S., Giroux, C.N., and Kleckner. N., Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family, *Cell*, 1997, vol. 88, pp. 375–384.
- Lander, N., Li, Z.-H., Niyogi, S., and Docampo. R., CRISPR/Cas9-induced disruption of paraflagellar rod protein 1 and 2 genes in *Trypanosoma cruzi* reveals their role in flagellar attachment, *mBio*, 2015, vol. 6. e01012.
- Li, Y. and Li. J., Bioanalytical chemistry based on CRISPR, *Progr. Chem.*, 2020, vol. 32, pp. 5–13.
- Limsirichai, P., Gaj, T., and Schaffer. D.V., CRISPR-mediated activation of latent HIV-1 expression, *Mol. Ther.*, 2016, vol. 24, pp. 499–507.
- Liu, J.-J., Orlova, N., Oakes, B.L., Ma, E., Spinner, H.B., Baney, K.L.M., Chuck, J., Tan, D., Knott, G.J., Harrington, L.B., Al-Shayeb, B., Wagner, A., Brötzmann, J., Staahl, B.T., Taylor, K.L., Desmarais, J., Nogales, E., and Doudna. J.A., CasX enzymes comprise a distinct family of RNA-guided genome editors, *Nature*, 2019, vol. 566, pp. 218–223.
- Maeder, M.L., Stefanidakis, M., Wilson, C.J., Baral, R., Barrera, L.A., Bounoutas, G.S., Bumcrot, D., Chao, H., Ciulla, D.M., DaSilva, J.A., Dass, A., Dhanapal, V., Fennell, T.J., Friedland, A.E., Giannoukos, G., Gloskowski, S.W., Glucksmann, A., Gotta, G.M., Jayaram, H., Haskett, S.J., Hopkins, B., Horng, J.E., Joshi, S., Marco, E., Mepani, R., Reyon, D., Ta, T., Tabbaa, D.G., Samuelsson, S.J., Shen, S., Skor, M.N., Stetkiewicz, P., Wang, T., Yudkoff, C., Myer, V.E., Albright, C.F., and Jiang. H., Development of a gene-editing approach to restore vision loss in Leber congenital amaurosis type 10, *Nat. Med.*, 2019, vol. 25, pp. 229–233.
- Makarova, K.S., Wolf, Y.I., Iranzo, J., Shmakov, S.A., Alkhnbashi, O.S., Brouns, S.J.J., Charpentier, E., Cheng, D., Haft, D.H., Horvath, P., Moineau, S., Mojica, F.J.M., Scott, D., Shah, S.A., Siksnyš, V., Terns, M.P., Venclovas, C., White, M.F., Yakunin, A.F., Yan, W., Zhang, F., Garrett, R.A., Backofen, R., van der Oost, J., Barrangou, R., and Koonin. E.V., Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants, *Nat. Rev. Microbiol.*, 2020, vol. 18, pp. 67–83.
- Mohamed, N.-V., Larroquette, F., Beitel, L.K., Fon, E.A., and Durcan. T.M., One step into the future: new iPSC tools to advance research in Parkinson's disease and neurological disorders, *J. Parkinsons Dis.*, 2019, vol. 9, pp. 265–281.
- Nagler, A., Vredevoogd, D.W., Alon, M., Cheng, P.F., Traibish, S., Kalaora, S., Arafah, R., Goldin, V., Levesque, M.P., Peeper, D.S., and Samuels. Y., A genome-wide CRISPR screen identifies FBXO42 involvement in resistance toward MEK inhibition in NRAS-mutant melanoma, *Pigment Cell Melanoma Res.*, 2020, vol. 33, pp. 334–344.
- Narimani, M., Sharifi, M., Hakhmaneshi, M.S., Roshani, D., Kazemi, M., Hejazi, S.H., and Jalili, A., *BIRC5* gene disruption via CRISPR/Cas9n platform suppress acute myelocytic leukemia progression, *Iran. Biomed. J.*, 2019, vol. 23, pp. 369–378.
- Nelles, D.A., Fang, M.Y., O'Connell, M.R., Xu, J.L., Markmiller, S.J., Doudna, J.A., and Yeo. G.W., Programmable RNA tracking in live cells with CRISPR/Cas9, *Cell*, 2016, vol. 165, pp. 488–496.
- Russell, T.A., Stefanovic, T., and Tschärke. D.C., Engineering herpes simplex viruses by infection-transfection methods including recombination site targeting by CRIS-

- PR/Cas9 nucleases, *J. Virol. Methods*, 2015, vol. 213, pp. 18–25.
- Sharma, P., Sharma, B.S., and Verma. R.J., CRISPR-based genome editing of zebrafish, *Prog. Mol. Biol. Transl. Sci.*, 2021, vol. 180, pp. 69–84.
- Shen, B., Brown, K.M., Lee, T.D., and Sibley. L.D., Efficient gene disruption in diverse strains of *Toxoplasma gondii* using CRISPR/CAS9, *mBio*, 2014, vol. 5. e01114-e01114.
- Tanihara, F., Hirata, M., and Otoi. T., Current status of the application of gene editing in pigs, *J. Reprod. Dev.*, 2021.
- Wei, C., Wang, F., Liu, W., Zhao, W., Yang, Y., Li, K., Xiao, L., and Shen. J., CRISPR/Cas9 targeting of the androgen receptor suppresses the growth of LNCaP human prostate cancer cells, *Mol. Med. Rep.*, 2018, vol. 17, pp. 2901–2906.
- Wu, Y., Liang, D., Wang, Y., Bai, M., Tang, W., Bao, S., Yan, Z., Li, D., and Li. J., Correction a of genetic disease in mouse via use of CRISPR–Cas9, *Cell Stem Cell*, 2013, vol. 13, pp. 659–662.
- Xu, A., Qin, C., Lang, Y., Wang, M., Lin, M., Li, C., Zhang, R., and Tang. J., A simple and rapid approach to manipulate pseudorabies virus genome by CRISPR/Cas9 system, *Biotechnol. Lett.*, 2015, vol. 37, pp. 1265–1272.
- Yang, H.-C., Lin, S.-R., Liu, C.-J., Kao, J.-H., Chen, D.-S., and Chen, P.-J., The CRISPR/Cas9 system facilitates elimination of the persistent intrahepatic HBV genomes in vivo, *J. Hepatol.*, 2014, vol. 60, p. 174.
- Zhang, H., Lu, X., Wang, Z., Yan, X., and Cui. H., Excretion from long glandular trichomes contributes to alleviation of cadmium toxicity in *Nicotiana tabacum*, *Environ. Pollut.*, 2021, vol. 285, p. 117184.
- Zhao, X., Liu, L., Lang, J., Cheng, K., Wang, Y., Li, X., Shi, J., Wang, Y., and Nie. G., A CRISPR-Cas13a system for efficient and specific therapeutic targeting of mutant KRAS for pancreatic cancer treatment, *Cancer Lett.*, 2018, vol. 431, pp. 171–181.