

# A Metric Approach to Hot Topics in Biomedicine via Keyword Co-occurrence

Jane H. Qin<sup>1,2</sup>, Jean J. Wang<sup>1,2</sup>, Fred Y. Ye<sup>1,2†</sup>

<sup>1</sup>Jiangsu Key Laboratory of Data Engineering and Knowledge Service, School of Information Management, Nanjing University, Nanjing 210023, China

<sup>2</sup>International Joint Informatics Laboratory (IJIL), Nanjing University – University of Illinois, Nanjing – Champaign, China – USA

Citation: Jane H. Qin, Jean J. Wang, Fred Y. Ye (2019). A metric approach to hot topics in biomedicine via keyword co-occurrence. *Journal of Data and Information Science*, 4(4), 13–25  
DOI: 10.2478/jdis-2019-0018  
Received: Sep. 24, 2019  
Revised: Nov. 19, 2019  
Accepted: Nov. 24, 2019

## Abstract

**Purpose:** To reveal the research hotspots and relationship among three research hot topics in biomedicine, namely CRISPR, iPS (induced Pluripotent Stem) cell and Synthetic biology.

**Design/methodology/approach:** We set up their keyword co-occurrence networks with using three indicators and information visualization for metric analysis.

**Findings:** The results reveal the main research hotspots in the three topics are different, but the overlapping keywords in the three topics indicate that they are mutually integrated and interacted each other.

**Research limitations:** All analyses use keywords, without any other forms.

**Practical implications:** We try to find the information distribution and structure of these three hot topics for revealing their research status and interactions, and for promoting biomedical developments.

**Originality/value:** We chose the core keywords in three research hot topics in biomedicine by using h-index.

**Keywords** Keyword co-occurrence; Network analysis; Information visualization; Biomedicine; Hot topics; CRISPR-Cas; iPS cell; Synthetic biology

## 1 Introduction

Since 2000, new century, biomedicine had great progress at both scientific and technical levels. Among “breakthrough of the year” in *Science* every year, we could conclude three important hot topics in biomedicine as follows. The first topic is Genome editing technique CRISPR/Cas (Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-associated system). As a genome editing method,



† Corresponding author: Fred Y. Ye (E-mail: yye@nju.edu.cn).

**Research Paper**

CRISPR/Cas was the top breakthrough in 2015. The CRISPR/Cas system is a prokaryotic immune system that confers resistance to foreign genetic elements such as those present within plasmids and phages that provides a form of acquired immunity. At beginning, CRISPR described segments of prokaryotic DNA containing short, repetitive base sequences in ancient bacteria (Horvath & Barrangou, 2010). Later, the group of Jennifer Doudna induced CRISPR/Cas9 as a tool to cut DNA with crRNAs in 2012 (Jinek et al.), and then the group of Feng Zhang applied CRISPR/Cas9 into eucaryotic cells in 2013 (Cong et al., 2013). The group of Ma et al. (2017) describe the correction of a pathogenic gene mutation in human embryos with CRISPR/Cas9. Cox et al. (2017) proved that RNA can be edited with CRISPR-Cas13 to correct disease-relevant human mutations and proposed an RNA-editing platform named REPAIR. While another nuclease Cpf1 was discovered in 2015 then CRISPR/Cpf1 became another CRISPR system (Zetsche et al., 2015). Yan et al. (2019) systematically discovered additional subtypes of type V CRISPR-Cas systems. The diversity, modularity, and efficacy of CRISPR-Cas systems are driving a biotechnological revolution and CRISPR-Cas guides the future of genetic engineering (Knott & Doudna, 2018). The second theme is Stem cell technique iPS cell. As a type of pluripotent stem cell, the iPS cell technique was selected into new breakthrough in both 2012 and 2016. The iPS cell technique was pioneered by Shinya Yamanaka's lab in Kyoto, who showed in 2006 that the introduction of four specific genes encoding transcription factors could convert adult cells into pluripotent stem cells (Takahashi, 2006), on which Yamanaka was awarded the 2012 Nobel Prize along with Sir John Gurdon for their discovery that mature cells can be reprogrammed to become pluripotent. Since then, researchers have found a variety of more optimal induction methods (Anokye-Danso et al., 2011; Ma, Kong, & Zhu, 2017). At the meantime, researchers turned to introduce disease-associated mutations into a sample of iPS cells through gene editing. Paquet et al. (2016) generated cells with precise combinations of Alzheimer's-associated mutations by introducing specific point mutations into iPS cells using CRISPR. The iPS cells have wide application perspectives in drug discovery and disease modelling (Scudellari, 2016). The last topic is Synthetic biology and artificial life. This is an interdisciplinary branch of biology and engineering, which was selected into new breakthrough in 2010. Synthetic biologists come in two broad classes. One uses unnatural molecules to reproduce emergent behaviors from natural biology, with the goal of creating artificial life. The other seeks interchangeable parts from natural biology to assemble into systems that function unnaturally (Benner & Sismour, 2005). Gibson et al. (2010) introduced their study about the Creation of a bacterial cell controlled by a chemically synthesized genome. Esvelt and Wang (2013) think Genome-modification technologies enable the rational engineering and perturbation



of biological systems, such as CRISPR/Cas. Cameron, Bashor and Collins (2014) reviews the history of synthetic biology and points out that the field of synthetic biology has chartered many notable achievements and is poised to transform biotechnology and medicine.

In this article, based on biomedical documents and data analysis, we try to find the information distribution and structure of these three hot topics via analyzing the collaborative networks and visualizing their cores, for revealing their research status and interactions, for promoting biomedical developments.

## 2 Methodology

We process core keyword co-occurrence networks in this study. The methods focus on network analysis (Friedkin, 1991; Newman, 2004; Wolfe, 1997) and information visualization (Chen, 2006), and data come from scientific document database. For information visualization, VOSviewer (Eck & Waltman, 2010) is applied to draw pictures. In network analysis, Gephi (Bastian, Heymann, & Jacomy, 2009) and UCINET (Borgatti, Everett, & Freeman, 2002) are applied to compute network parameters. Meanwhile, open software SATI (Liu & Ye, 2012), Excel and R programming are used for data processing.

### 2.1 Methods

In this article, we chose the core keywords in three research hot topics in biomedicine by using h-index. Hirsch (2005) proposed the h-index, defined as the number of papers with citation number  $\geq h$ , as a useful index to characterize the scientific output of a researcher. If all papers published by one research are arranged in descending order of citation frequency, supposing  $c_i$  is the total number of citations of the  $i$ th paper, h-index can be quantified by formula  $h = \{\max i: i \leq c_i\}$ .

Because h-index takes into account both the number and quality of papers published by one researcher, it overcomes the shortcomings of the previous single dimension theory, such as the number of papers or the number of citations (Bornmann & Daniel, 2005). Therefore, h-index can more objectively evaluate the academic achievements of researchers, which has brought widespread attention in scientific circles (Bornmann & Daniel, 2007). Braun, Glänzel, and Schubert (2006) put forward that h-index can be applied in the evaluation of impact of journals for the first time, and believed that h-index was a powerful supplement to the impact factor of journals. Banks (2006) applied h-index to identify the main research topics of compounds. Thereafter, h-index is used in many academic research areas. Its definition is extended as follows: h-index of one academic information source refers to that at least h articles has been cited at least h times (Ye, 2014).



Keyword co-occurrence was considered to be the main means for identifying research themes (Wang et al., 2017). Keyword co-occurrence network reflects the knowledge structure and knowledge kernel that can display the relationship between keywords (Su & Lee, 2010). Lee and Su (2010) believed that research hotspots can be evaluated by the centrality of the nodes in a keyword co-occurrence network. In keyword co-occurrence network, nodes represent keywords, while edges represent co-occurrence relationships among nodes. By using the social network analysis method to analyze the keyword co-occurrence networks, we can analyze the knowledge structure and hotspots of the research field.

This article evaluates the influence of nodes in the network based on closeness centrality, betweenness centrality and eigenvector centrality. Degree centrality means the importance of nodes in the network. The higher degree centrality of the node, the more important the node is, which means that the keywords represented by the node are more likely to be research hotspots. Betweenness centrality measures the ability of one keyword in a network to affect the other keywords that appear together. eigenvector centrality measures the number of adjacency nodes and the influence of adjacency nodes.

### (1) Degree Centrality

In network measures, centrality indices are applied as terms of a real-valued function on the vertices of a graph, where the values produced are expected to provide a ranking which identifies the most important nodes. For a given graph  $G(V, E)$  with number of vertices  $V$  and number of edges  $E$ , let  $A=(a_{u,v})$  be the adjacency matrix, i.e.  $a_{u,v} = 1$  if vertex  $u$  is linked to vertex  $v$  and  $a_{u,v} = 0$  otherwise. The degree centrality score of vertex  $u$  can be defined as

$$x_u = \sum_{v \in G} a_{u,v} x_v \quad (1)$$

The relative centrality of vertex  $u$  can be defined as

$$x_u = \frac{1}{\lambda} \sum_{v \in G} a_{u,v} x_v \quad (2)$$

### (2) Betweenness Centrality

Betweenness Centrality measures the shortest path in a network, which is used to evaluate the role of nodes in information integration in social networks. The higher the betweenness centrality, the greater the role it plays in information integration.  $G_{st}$  represents the number of shortest paths from point  $s$  to point  $t$ .  $G_{st}(v)$  represents the number of shortest paths from point  $s$  to point  $t$  that pass through node  $v$ . The betweenness centrality of vertex  $u$  can be defined as follows:

$$x_u = \sum_{v \in G} \frac{G_{st}(v)}{G_{st}} \quad (3)$$



### (3) Eigenvector Centrality

Since the entries in the adjacency matrix are non-negative, there is a unique largest eigenvalue, which is real and positive. This greatest eigenvalue results in the desired centrality measure is eigenvector centrality or eigencentality, which reveals the core importance of a vertex in a network. Its eigenvectors are orthogonal and diagonalizable. The centrality of vertices is proportional to the sum of the central points of the vertices it connects. The eigenvector center  $x$  is described in two equivalent ways. As the sum of matrix equations, the eigenvector centrality can be defined as follows:

$$Ax = \lambda x \quad (4)$$

$$Ax = \lambda x, \lambda x_i = \sum_{j=1}^n a_{ij} x_j, i = 1, \dots, n \quad (5)$$

$\lambda$  is the maximum eigenvalue of  $A$  and  $n$  is the number of vertices.

## 2.2 Data

In our empirical study we search the Web of Science (WoS) database for articles published during 1900 to 2018. We collected the data in January 2019. The retrieval strategies were as follows:

### (1) H1-CRISPR

TS="clustered regularly interspaced short palindromic repeats" OR CRISPR

### (2) H2-iPS cell

TS="induce\* pluripotent stem cell" OR "induce\* pluripotent stem cells" OR "IPS cell" OR "IPS cells"

### (3) H3-Synthetic biology

TS="synthetic biology" OR "gene circuit" OR "gene circuits" OR "genetic circuit" OR "genetic circuits" OR "genetic device" OR "genetic devices" OR "synthetic life" OR "synthetic lives" OR "synthetic tissue" OR "synthetic tissues" OR "synthetic cell" OR "synthetic cells" OR "synthetic genome" OR "synthetic genomes" OR "synthetic gene" OR "synthetic genes" OR "minimal genome" OR "minimal genomes" OR "biology, synthetic"

The computed data will lead to next results for finding core keywords and setting up keyword co-occurrence networks.

## 3 Keyword co-occurrence results

High-frequency keywords can reflect the research hotspots and research directions to some extent, but the information displayed by the linear arrangement of the



frequency of keywords has great limitations. H-index can comprehensively reflect the occurrence frequency of keywords and the number of citations. In this article, the 100 keywords with the highest h-index are defined as core keywords. Keyword co-occurrence relationship can reflect the internal connection between keywords. In this chapter, we construct the co-occurrence matrix and co-word network with the help of R language. Besides that, we analyze the research hotspots of the three hot topics in biomedicine by using social network analysis method.

In this study, R language is used to extract keywords and the number of citations corresponding to keywords. The h-index is calculated by our own programming. The extracted keywords have the following problems.

- (1) Case difference, such as “Induced pluripotent stem cell”, “induced pluripotent stem cell”, “CRISPR/Cas9”, and “CRISPR/Cas9”.
- (2) Inconsistent connectors, such as “CRISPR/Cas9”, and “CRISPR-Cas9”.
- (3) Heteronyms, such as abbreviation and the full name phenomenon. “iPSCs”, “Human-induced pluripotent stem cells”, and “Induced pluripotent stem cells (iPSCs)” have the same meaning.

Considering that many keywords are special terms in the field of biomedicine, and the current general dictionary is not applicable for the study. The following two methods are used for data processing. The first method is case conversion, which unifies keywords into capital letters. The second method is self-compiled dictionary which can solve the problems of inconsistent connectors and Heteronyms.

### 3.1 Keyword co-occurrence with CRISPR/Cas9

Fig. 1 shows the co-occurrence network of 100 keywords with the highest h-index in the field of CRISPR/Cas9. Node color represents matrix-based clustering and node size represents h-index. Table 1 shows the 20 nodes with the highest degree centrality and betweenness centrality.

In the network, “CRISPR”, “GENOME EDITING”, “CRISPR/CAS”, “GENOME ENGINEERING”, “HOMOLOGOUS RECOMBINATION”, “GENE TARGET” are located in the center of the network, and have high degree centrality and betweenness centrality. Thus, they are the core research contents. “ZEBRAFISH”, “MOUSE”, “ZFN”, “TALEN”, “GENE THERAPY”, “CANCER” are the much important research.

Interestingly, the keywords of “IPSC”, “HUMAN IPSC”, “STEM CELL”, “SYNTHETIC BIOLOGY”, “METABOLIC ENGINEERING” are conspicuous in the co-occurrence network. Besides that, these keywords have high degree centrality and betweenness centrality. That means CRISPR/Cas9 and the other two hot topics have large cross-study.



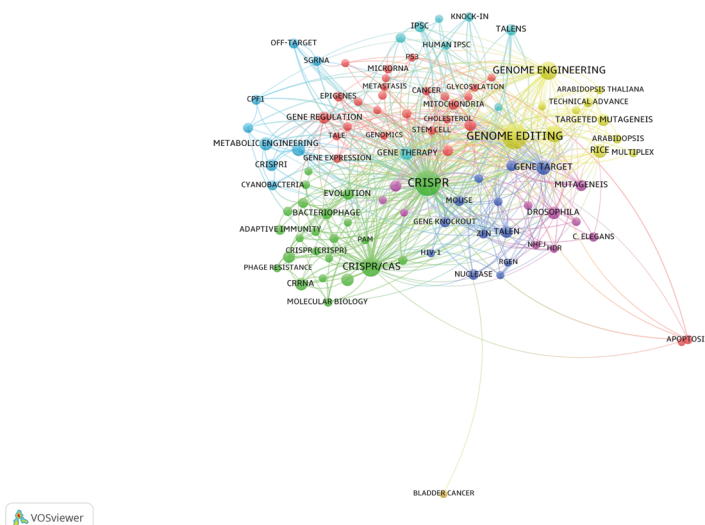


Figure 1. Co-occurrence network with the highest h-index keywords in the field of CRISPR/Cas9.

Table 1. Co-occurrence network centrality with the highest h-index keywords in the field of CRISPR/Cas9.

Rank	Keyword	Degree Centrality	Keyword	Betweenness Centrality
1	CRISPR	188	CRISPR	1,378.00
2	GENOME EDITING	156	GENOME EDITING	664.00
3	CRISPR/CAS	132	CRISPR/CAS	441.37
4	GENOME ENGINEERING	94	GENOME ENGINEERING	157.52
5	HOMOLOGOUS RECOMBINATION	84	GENES	135.44
6	GENES	76	HOMOLOGOUS RECOMBINATION	111.44
7	ZEBRAFISH	70	ZEBRAFISH	110.97
8	GENE TARGET	66	GENE REGULATION	60.97
9	TALEN	64	CRISPRI	53.39
10	SYNTHETIC BIOLOGY	58	APOPTOSIS	49.33
11	GENE REGULATION	56	DNA REPAIR	48.34
12	GENE THERAPY	56	SYNTHETIC BIOLOGY	47.97
13	DNA REPAIR	54	GENE TARGET	47.13
14	GENE KNOCKOUT	54	IPSC	42.64
15	IPSC	50	GENE KNOCKOUT	41.15
16	ZFN	50	TALEN	41.09
17	SGRNA	46	GENE THERAPY	39.40
18	CANCER	42	EVOLUTION	37.89
19	CRRNA	42	SGRNA	34.15
20	EVOLUTION	42	CANCER	34.14

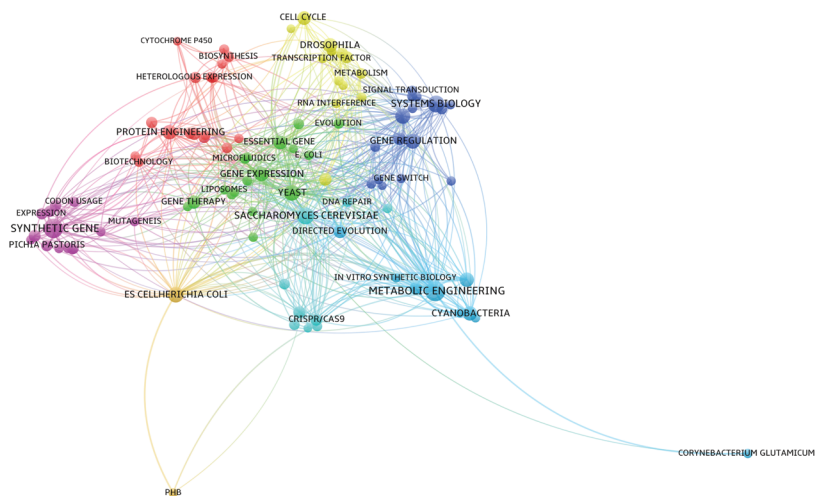






Table 2. Co-occurrence network centrality with the highest h-index keywords in the field of iPS cell.

Rank	Keyword	Degree Centrality	Keyword	Betweenness Centrality
1	STEM CELL	184	STEM CELL	359.39
2	EMBRYONIC STEM CELL	170	EMBRYONIC STEM CELL	267.07
3	REPROGRAMMING	166	HUMAN IPSC	254.20
4	HUMAN IPSC	162	REPROGRAMMING	246.40
5	PLURIPOTENT STEM CELL	160	DIFFERENTIATION	225.41
6	DIFFERENTIATION	158	PLURIPOTENT STEM CELL	221.89
7	REGENERATIVE MEDICINE	118	HUMAN EMBRYONIC STEM CELL	110.40
8	HUMAN EMBRYONIC STEM CELL	116	REGENERATIVE MEDICINE	100.92
9	NEURAL STEM CELL	116	MESENCHYMAL STEM CELL	94.19
10	MESENCHYMAL STEM CELL	114	NEURAL STEM CELL	83.66
11	CELL THERAPY	102	TRANSPLANTATION	70.11
12	PLURIPOTENCY	102	PLURIPOTENCY	67.90
13	TRANSPLANTATION	102	CELL THERAPY	60.74
14	TISSUE ENGINEERING	98	CARDIOMYOCYTES	58.25
15	CARDIOMYOCYTES	90	NEURON	56.58
16	NEURON	90	TISSUE ENGINEERING	54.92
17	DISEASE MODELING	82	HIPSC	43.86
18	HIPSC	82	DRUG SCREENING	42.39
19	PARKINSON'S DISEASE	82	GENE EXPRESSION	40.25
20	GENE EXPRESSION	80	DISEASE MODELING	37.50



VOSviewer

Figure 3. Co-occurrence network with the highest h-index keywords in the field of synthetic biology.



centrality. Thus, they are the core research contents. “YEAST”, “ESCELLHERICHIA COLI” and “SACCHAROMYCES CEREVISIAE” are experimental vectors used in synthetic biology research. These keywords are located in the network center and have high degree centrality and betweenness centrality, which indicates that they are widely used in research .

In the keyword co-occurrence network, CRISPR/CAS9 also have high degree centrality, and they are important node in the keyword cluster. This indicates that there are correlation between synthetic biology and CRISPR/CAS9. Besides that, the frequency of cross research and the number of citations is relatively high.

Table 3. Co-occurrence network centrality with the highest h-index keywords in the field of synthetic biology.

Rank	Keyword	Degree Centrality	Keyword	Betweenness Centrality
1	METABOLIC ENGINEERING	100	METABOLIC ENGINEERING	378.42
2	YEAST	90	GENE EXPRESSION	312.93
3	GENE CIRCUIT	82	YEAST	286.57
4	ES CELLHERICHIA COLI	80	GENE CIRCUIT	281.48
5	GENE EXPRESSION	80	ES CELLHERICHIA COLI	249.28
6	SACCHAROMYCES CEREVISIAE	80	SACCHAROMYCES CEREVISIAE	229.17
7	PROTEIN ENGINEERING	66	SYNTHETIC GENE	180.28
8	DIRECTED EVOLUTION	62	PROTEIN ENGINEERING	171.01
9	GENE REGULATION	60	CELL CYCLE	169.32
10	SYNTHETIC GENE	60	GENE THERAPY	158.88
11	SYSTEMS BIOLOGY	60	TRANSCRIPTION	146.23
12	TRANSCRIPTION	60	DIRECTED EVOLUTION	143.33
13	GENE ENGINEERING	56	SYSTEMS BIOLOGY	130.95
14	BIOTECHNOLOGY	50	GENE REGULATION	108.56
15	GENE THERAPY	50	GENE ENGINEERING	90.24
16	CRISPR/CAS9	48	ESSENTIAL GENE	87.67
17	CYANOBACTERIA	46	CELL-FREE PROTEIN SYNTHESIS	76.44
18	ESSENTIAL GENE	46	EVOLUTION	71.90
19	EVOLUTION	44	TRANSCRIPTION FACTOR	70.90
20	E. COLI	42	BIOTECHNOLOGY	66.57

#### 4 Discussion and conclusion

Above results construct core keyword co-occurrence networks with visualizing and calculating the keywords' centralities. The three research hot topics in biomedicine are analyzed and characterized as follows.

- (1) The research hotspots of CRISPR/Cas9 include the comparison of gene editing technology with the previous two generations, the discovery of new CRISPR/Cas9 system, improvement of gene editing technology and methods, and application research of CRISPR/Cas9 in the gene therapy and cancer therapy.



- (2) The research hotspots of synthetic biology include “METABOLIC ENGINEERING”, “GENE CIRCUIT”, “GENE EXPRESSION”, “SYSTEMS BIOLOGY”, “SYSTEMS BIOLOGY”, and “PROTEIN ENGINEERING”.
- (3) The research hotspots of iPS cells include HUMAN IPSC, the comparison between iPS cell and EMBRYONIC STEM CELL, and the application of iPS cells in the research of CARDIOMYOCYTES, NEURON, PARKINSON’S DISEASE, etc.
- (4) There were overlapping keywords corresponding to the three biomedical topics, among which the overlapping keywords of synthetic biology and CRISPR/Cas9 were the most obvious. The research on the three topics is overlapping.

Since all analyses use keywords, without any other forms, the limitations are remained in this article, which may be improved in future studies.

## Acknowledgements

We acknowledge the National Natural Science Foundation of China Grant 71673131 for financial support.

## Author Contributions

Jane J. Qin (jhqin@smail.nju.edu.cn) collected and processed the data and wrote Chinese version, Jean J. Wang (wangjingjing236@sina.com) re-organized and wrote English version, and Fred Y. Ye (yye@nju.edu.cn) initiated the research and revised the paper.

## Reference

- Anokye-Danso, F., Trivedi, C.M., Juhr, D., Gupta, M., Cui, Z., Tian, Y., Zhang, Y., Yang, W., Gruber, P.J., & Epstein, J.A. (2011). Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency. *Cell Stem Cell*, 8(4), 376–388.
- Banks, M.G. (2006). An extension of the hirsch index: Indexing scientific topics and compounds. *Scientometrics*, 69(1), 161–168.
- Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. In *Proceedings of the Third International Conference on Weblogs and Social Media* (361–362). San Jose, California, USA. ICWSM.
- Benner, S.A., & Sismour, A.M. (2005). Synthetic biology. *Nat Rev Genet*, 6, 533–543.
- Borgatti, S.P., Everett, M.G., & Freeman, L.C. (2002). *Ucinet for windows: Software for social network analysis*, Harvard, MA: Analytic Technologies.
- Bornmann, L., & Daniel, H.D. (2005). Does the h-index for ranking of scientists really work? *Scientometrics*, 65(3), 391–392.
- Bornmann, L., & Daniel, H. (2007). What do we know about the h index? *Journal of the American Society for Information Science and Technology*, 58(9), 1381–1385.



**Research Paper**

- Braun, T., Glänzel, W., & Schubert, A. (2006). A hirsch-type index for journals. *Scientometrics*, 69(1), 169–173.
- Cameron, D.E., Bashor, C.J., & Collins, J.J. (2014). A brief history of synthetic biology. *Nature Reviews Microbiology*, 12, 381–390.
- Chen, C. (2006). *Information visualization: Beyond the horizon*, Berlin: Springer.
- Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., & Marraffini, L.A. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science*, 339, 819–823.
- Cox, D.B.T., Gootenberg, J.S., Abudayyeh, O.O., Franklin, B., Kellner, M.J., Joung, J., & Zhang, F. (2017). Rna editing with crispr-cas13. *Science*, 358, 1019–1027.
- Eck, N.J.V., & Waltman, L. (2010). Software survey: Vosviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2), 523–538.
- Esvelt, K.M., & Wang, H.H. (2013). Genome-scale engineering for systems and synthetic biology. *Molecular Systems Biology*, 9, 641–641.
- Friedkin, N.E. (1991). Theoretical foundations for centrality measures. *American Journal of Sociology*, 96, 1478–1504.
- Gibson, D.G., Glass, J.I., Lartigue, C., Noskov, V.N., Chuang, R.Y., Algire, M.A., Benders, G.A., Montague, M.G., Ma, L., & Moodie, M.M. (2010). Creation of a bacterial cell controlled by a chemically synthesized genome. *Science*, 329(5987), 52–56.
- Hirsch, J.E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 16569–16572.
- Horvath, P., & Barrangou, R. (2010). CRISPR/Cas, the immune system of bacteria and archaea. *Science*, 327, 167–170.
- Knott, G.A., & Doudna, J.A. (2018). CRISPR-Cas guides the future of genetic engineering. *Science*, 361, 866–869.
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., & Charpentier, E. (2012). A programmable dual-rna-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 337, 816–821.
- Lee, P.C., & Su, H.N. (2010). Investigating the structure of regional innovation system research through keyword co-occurrence and social network analysis. *Innovation: Management, Policy & Practice*, 12(1), 26–40.
- Liu, Q., & Ye, Y. (2012). A study on mining bibliographic records by designed software: a case study on library and information science. *Journal of Information Resources Management*, 22(1), 50–58.
- Ma, H., Marti-Gutierrez, N., Park, S.W., Wu, J., & Mitalipov, S. (2017). Correction of a pathogenic gene mutation in human embryos. *Nature*, 548, 413–419.
- Ma, X., Kong, L., & Zhu, S. (2017). Reprogramming cell fates by small molecules. *Protein Cell*, 8(5), 328–348.
- Newman, M. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Science*, 101, 5200–5205.
- Paquet, D., Kwart, D., Chen, A., Sproul, A., Jacob, S., Teo, S., Olsen, K.M., Gregg, A., Noggle, S., & Tessier-Lavigne, M. (2016). Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. *Nature*, 533, 125–129.



- Scudellari, M. (2016). A decade of ips cells. *Nature*, 534, 310–312.
- Su, H.N., & Lee, P.-C. (2010). Mapping knowledge structure by keyword co-occurrence: A first look at journal papers in technology foresight. *Scientometrics*, 85(1), 65–79.
- Takahashi, K. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, 126, 663–676.
- Wang, P., Zhu, F., Song, H., & Hou, J. (2017). A bibliometric profile of current science between 1961 and 2015. *Current Science*, 113(3), 386–392.
- Wolfe, A.W. (1997). Social network analysis: Methods and applications. *American Ethnologist*, 24, 136–137.
- Yan, W.X., Hunnewell, P., Alfonse, L.E., Carte, J.M., Keston-Smith, E., Sothiselvam, S., Garrity, A.J., Chong, S., Makarova, K.S., Koonin, E.V., Cheng, D.R., & Scott, D.A. (2019). Functionally diverse type v CRISPR-Cassystems. *Science*, 363(6422), 88–91.
- Ye, F.Y. (2014). The research progress and developing perspective of assessment indicators. *Journal of the China Society for Scientific and Technical Information*, 33, 215–224.
- Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., Van Der Oost, J., Regev, A., V.Koonin, E., & Zhang, F. (2015). Cpf1 is a single rna-guided endonuclease of a class 2 CRISPR-Cassystem. *Cell*, 163, 759–771.



This is an open access article licensed under the Creative Commons Attribution-NonCommercial-NoDerivs License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

