Overview of Trends in Global Single Cell Research Based on Bibliometric Analysis and LDA Model (2009–2019)

Tian Jiang¹, Xiaoping Liu¹, Chao Zhang¹, Chuanhao Yin², Huizhou Liu^{1†}

¹National Science Library, Chinese Academy of Sciences, Beijing 100190, China

²Chinese Institute of Electronics, Beijing 100036, China

Citation: Jiang, T., Liu, X.P., Zhang, C., Yin, C.H., & Liu, H.Z. (2021). Overview of trends in global single cell research based on bibliometric analysis and LDA model (2009–2019). *Journal of Data and Information Science*, 6(2), 163–178. <https://doi.org/10.2478/jdis-2021-0008>

Received: Sep. 8, 2020
Revised: Oct. 11, 2020
Accepted: Oct. 23, 2020

Abstract

Purpose: This article aims to describe the global research profile and the development trends of single cell research from the perspective of bibliometric analysis and semantic mining.

Design/methodology/approach: The literatures on single cell research were extracted from Clarivate Analytic's Web of Science Core Collection between 2009 and 2019. Firstly, bibliometric analyses were performed with Thomson Data Analyzer (TDA). Secondly, topic identification and evolution trends of single cell research was conducted through the LDA topic model. Thirdly, taking the post-discretized method which is used for topic evolution analysis for reference, the topics were also be dispersed to countries to detect the spatial distribution.

Findings: The publication of single cell research shows significantly increasing tendency in the last decade. The topics of single cell research field can be divided into three categories, which respectively refers to single cell research methods, mechanism of biological process, and clinical application of single cell technologies. The different trends of these categories indicate that technological innovation drives the development of applied research. The continuous and rapid growth of the topic strength in the field of *cancer diagnosis and treatment* indicates that this research topic has received extensive attention in recent years. The topic distributions of some countries are relatively balanced, while for the other countries, several topics show significant superiority.

Research limitations: The analyzed data of this study only contain those were included in the Web of Science Core Collection.

Practical implications: This study provides insights into the research progress regarding single cell field and identifies the most concerned topics which reflect potential opportunities and challenges. The national topic distribution analysis based on the post-discretized analysis method extends topic analysis from time dimension to space dimension.

† Corresponding author: Huizhou Liu (E-mail: liuhz@mail.las.ac.cn; liuhz_las@outlook.com).

Research Paper

Originality/value: This paper combines bibliometric analysis and LDA model to analyze the evolution trends of single cell research field. The method of extending post-discretized analysis from time dimension to space dimension is distinctive and insightful.

Keywords LDA model; Topic evolution; Bibliometric analysis; Post-discretized; Single-cell

1 Introduction

Traditional cell biology analyses are performed using bulk cells, which mask the difference among cells. Research on the single cell level can dissect the cell-to-cell variation and heterogeneity, which provides powerful means to reveal the mechanism of cell fate decision, embryogenesis and organogenesis, and also provides new methods for tumor targeted therapy (Briggs et al., 2018; Cao et al., 2019; Farrell et al., 2018; Griffiths, Scialdone, & Marioni, 2018; Wei et al., 2016). In recent years, the rapid development of a variety of single cell technology has promoted the continuous deepening of single cell research. Because individual cell may occur in different microenvironment or different stages of cell cycle, even gene expression of pure cell types are heterogeneous (Junker & van Oudenaarden, 2014). High throughput single cell sequencing technology including single cell genomic sequencing and single cell RNA-seq (scRNA-seq) can detect the gene structure and expression of individual cells, which is of great significance to the mapping of cells and the diagnosis and treatment of tumors. Zhang et al. (2018) mapped t-cell immunity in lung cancer and colorectal cancer on single cell level, revealed the subgroup classification, tissue distribution characteristics, intra-tumor population heterogeneity and drug target gene expression of lung cancer and colorectal cancer T cells, which is very important to the diagnosis and treatment of lung cancer and colorectal cancer (Guo et al., 2018). The mechanism of cell state and cell fate decision have always been common concerns in the process of organ development. Two kinds of single cell ChIP-seq technology with widely applicable and simple operation styles have been developed recently, which can adapt to different research needs, and analyze the mechanism of cell fate decision under development and disease conditions (Ai et al., 2019; Wang et al., 2019). In addition, microfluidic chip, flow cytometry, single cell living imaging, and other techniques also play important roles in single cell research (Lindström, 2012; Reece et al., 2016).

Bibliometric is an effective tool for quantitative analysis of scientific and technological literature (Nicolaisen, 2010), which is widely used to evaluate research trends in many fields (Zhang et al., 2016; Zheng et al., 2016). As its rising importance in life science and clinical application, the profound investigation in literature

of single cell research is paramount. To the best of our knowledge, there is no bibliometric analysis in the single cell research field until this study.

Topic model serves as an effective tool for text mining in the science and technological papers, which can identify the research topics and hotspots. Latent Dirichlet Allocation (LDA) is one of the most popular topic model which has been applied in various fields (Jelodar et al., 2019).

In this paper, we combine the bibliometric method and LDA model to analyze the development trend of single cell research from the perspective of both statistical analysis and text mining. Besides, taking the post-discretized method for reference, the topics were dispersed to the top10 productive countries to detect the topic distributions of these countries.

2 Research methodology

2.1 Data collection and bibliometric analysis

The data were drawn from Clarivate Analytics' Web of Science (WoS) Core Collection in April 2020. We used the following search query: TS=((“single cell*”) NOT (“fuel-cell*” or “membrane-fuel-cell*” or “oxide-fuel-cell*” or “yuannhsolid-oxide-fuel-cell*” or “SOFC” or “proton-exchange-membrane-fuel-cell*” or “PEMFC” or “Direct-methanol-fuel-cell*” or “vanadium-redox-flow-battery*” or “solar-cell*” or “membrane-electrode-assembly” or “electrocatalyst” or “electrolyte” or “oxygen-reduction-reaction” or “reactive-oxygen-species” or “electrode” or “cathode” or “anode” or “electric-field” or “conductivity” or “durability” or “electrochemistry” or “electrochemical-performance” or “electrooxidation” or “Impedance” or “Impedance spectroscopy” or “Polymer-electrolyte-membrane” or “graphene” or “algorithm”)). The study was restricted to peer reviewed research papers (articles and reviews) between 2009 and 2019. Meeting abstracts, proceedings papers, notes, corrections, editorial material, and letters were thus excluded. A total of 30,804 publications were collected.

Thomson Data Analyzer (TDA) and Microsoft Excel were employed for bibliometric study. Gephi was utilized for national collaboration analysis. The size of nodes represents the number of publications and the size of the lines represents the frequency of collaboration.

2.2 Topic identification and evolution analysis

2.2.1 Topic modeling

LDA model is used for topic detection in this paper, which is a three-layer Bayesian model proposed by Blei, Ng, and Jordan (2003). It is based on the idea that a document is represented as a random mixture over latent topics and a topic

Research Paper

is characterized by a distribution of words (Jelodar et al., 2019). The topics can be interpreted through the distribution of words with probabilities arranged in descending order. In this study, titles, abstracts, and keywords were extracted to form the corpus for LDA analysis. The topic modeling process is conducted with Python package called Gensim (<https://radimrehurek.com/gensim/models/ldamodel.html>). We ran the algorithm with the following settings: $\alpha = 50 / \text{num_top}$, $\beta = 0.1$, iterations = 1000, passes = 20. The top30 keywords were chosen to interpret the identified topics.

2.2.2 Determination of topic number

The determination of topic number directly affects the topic recognition by LDA model. Perplexity and the average similarity of topics are considered to determine the optimal number of topics. Perplexity is an index to evaluate the language model (Blei, Ng, & Jordan, 2003), its calculation formula is as follows:

$$\text{Perplexity}(D) = \exp \left\{ - \frac{\sum_{d=1}^M \log p(w_d)}{\sum_{d=1}^M N_d} \right\} \quad (1)$$

D represents the test set of corpus, M represents the number of documents in the test set, N_d represents the number of words in the d document, and $p(w_d)$ represents the probability distribution of words in the d document. Generally, with the increasing of the number of topics, perplexity tends to decline. The lower the degree of perplexity, the stronger the generalization ability of LDA model.

The average similarity of topics is an index to measure the average degree of difference among all topics, which is usually measured based on Jensen-Shannon divergence (JS divergence) (Lee, 2001), its calculation formula is as follows:

$$\text{avg_sim}(T_i, T_j) = \frac{\sum_{i=1}^{k-1} \sum_{j=i+1}^k JS(T_i \| T_j)}{K \times (K-1) / 2} \quad (2)$$

T_i and T_j represent two topics respectively, and $JS(T_i \| T_j)$ represents the JS divergence between topic T_i and T_j . Generally, the average similarity of topics increases with the increasing of topic numbers, and the possibility of recurring topics also increases.

According to the Perplexity-average similarity curve, the position where perplexity decreases gradually and the value of average similarity is relatively small tend to be selected as the optimal number of topics.

2.2.3 Topic evolution analysis

The post-discretized analysis is a method to perform topic evolution trends based on LDA model. Firstly, topics are identified on the whole data set through LDA

model. Secondly, the topics are discretized to different periods according to the time information. Topic evolution trends can be obtained by analyzing topic strengths in different periods. Topic strength describes the degree to which a topic receives attention in a certain time window. It can be expressed as the ratio of the total weight of the research topic in all documents to the total number of documents. Suppose θ_z^d is the proportion of topic z in document d , D_t is the text set on time window t , then the intensity of theme z on time window t can be expressed as: $\theta_z^t = \frac{\sum_{d=1}^{D_t} \theta_z^d}{D_t}$.

3 Results

3.1 Bibliometric analysis

3.1.1 Publication analysis

Bibliometric analysis of single-cell research field was performed by TDA. The publications of single cell research rose significantly in the last decade and it shows a more prominent rising tendency in the years to come, as can be seen in Figure 1.

Figure 1. Annual publications on single cell research from 2009 to 2019 (Based on the WoS data).

3.1.2 Country analysis

Figure 2a shows the ten countries with the highest number of publications on single cell research. All of them are developed countries except for China. Among these countries, the United States produced the highest amount of publications (12,556 articles: 40.76% of the total), much more than that of China (4,132 articles) which ranked second. Besides, the United States is also the country with the highest number of total citations, indicating that it takes the leading position in single cell

Research Paper

research. As for the number of per paper citations, Switzerland takes the first place with 32.61. Netherlands ranks fourth with 30.26, although it has the least publications among the 10 countries. China and Japan are the only two countries with the number of per paper citations below 20.00 despite their publications rank second and fifth respectively.

A collaboration network for top10 productive countries is shown in Figure 2b. The United States has most collaborations with the other countries. The US-China collaboration ranks first with 971 collaborated papers, followed by the US-Germany and the US-UK collaborations with 701 and 689 collaborated papers respectively.

Figure 2. Production and collaboration analysis of countries. a. Top10 most productive countries. b. Collaboration network of the top10 most productive countries.

3.2 Topic identification

According to the Perplexity-avg_sim curve and topic identification effect, K=20 was chosen as the optimum topic number (Figure 3). The top30 frequent terms with highest probability in the topic-keywords distributions were chosen to interpret the identified topics. After filtration of insignificant and repetitive topics, nineteen topics were identified. Due to the space limitation, only the top20 frequent terms of each topic are displayed in Table 1.

These topics can be divided into three categories. The first one is about single cell research methods, which include topic2, 7, 10, and 17. The second one is research on mechanism of biological process on single cell level, which include topic0, 1, 4, 5, 6, 8, 9, 11, 12, and 18. The third one is about single cell research application, which include topic3, 13, 14, 15, and 16.

Figure 3. Perplexity-avg_sim curve of LDA model. *avg_sim*: average similarity.

Table 1. Topics identified by LDA model (K=20).

Topic	Potential topic	Top20 frequent terms
0	Pathology of brain disease	disease; brain; distribution; relationship; region; pattern; nucleus; age; immunohistochemistry; human brain; cortex; proportion; purpose; hippocampus; cluster; focus; rat; input; neuron; pathology
1	Mathematical modeling of cell cycle	model; type; control; mechanism; betum; datum; network; cell cycle; complexity; framework; phase; association; simulation; modeling; mouse model; differential expression; account; interplay; prediction; balance
2	Single cell detection platform	detection; imaging; platform; fluorescence; sensitivity; design; resolution; quantification; device; mass spectrometry; living; sample; measurement; spectrometry; flow; array; magnitude; capability; microfluidic device; chip
3	Immune response	single-cell level; flow cytometry; infection; biology; single-cell analysis; cytometry; host; memory; emergence; virus; pathogenesis; set; health; inflammation; immune response; immune system; complex; initiation; immunity; site
4	Signal transduction	response; activity; vivo; activation; pathway; receptor; factor; target; inhibition; phenotype; replication; inhibitor; signaling; overexpression; depletion; miRNA; kinase; zebrafish; oxygen; angiogenesis
5	Phylogeny on single cell level	sequencing; identification; protocol; diversity; genome; situ hybridization; selection; mutation; life; amplification; analysis; sequence; accuracy; family; transfer; chromosome; classification; genus; total; syndrome
6	Intracellular calcium modulation	combination; generation; increase; frequency; alpha; action; calcium; injury; channel; layer; heart; stability; difference; modulation; transmission; strength; central nervous system; ion; mu m; mechanism
7	Single cell gel electrophoresis	treatment; damage; assay; exposure; repair; stress; comparison; apoptosis; cell line; extent; risk; carcinoma; peripheral blood; kidney; glioblastoma; liver; assessment; evaluation; radiation; toxicity
8	Molecular mechanism of embryonic development	development; mouse; tissue; single-cell resolution; embryo; transcriptome; origin; stage; cell type; mapping; morphogenesis; establishment; skin; immunofluorescence; nervous system; gap; epithelium; molecular mechanism; specification; resource

Research Paper

Table 1. Continued

Topic	Potential topic	Top20 frequent terms
9	Cell adhesion	microscopy; surface; interaction; adhesion; spectroscopy; different cell; manipulation; motility; extracellular matrix; binding; chemical; cell surface; force; substrate; bacterium; atomic force; speed; aeruginosa; spectra; spectrum
10	Isolation and sorting of single cell	single cell; range; isolation; quality; cycle; engineering; antibody; viability; delivery; screening; enrichment; field; high throughput; throughput; amount; droplet; suspension; cell viability; red blood; solution
11	Cell migration	protein; migration; rate; contrast; loss; context; absence; transition; invasion; cell division; localization; density; organism; division; cell migration; cell size; decrease; fraction; literature; core
12	Cell-to-cell variability analysis	expression; gene expression; gene; population; transcription; regulation; mRNA; evolution; variability; variation; chromatin; transcription factor; protein expression; correlation; noise; phenotypic; promoter; reporter; cell-to-cell variability; gene regulation
13	Cancer diagnosis and treatment	cancer; tumor; blood; resistance; therapy; progression; survival; breast cancer; drug; patient; metastasis; death; lung; efficacy; diagnosis; treatment; melanoma; microenvironment; cell death; persistence
14	Single cell oil	growth; production; yeast; concentration; metabolism; ratio; composition; content; reduction; source; accumulation; plant; degradation; synthesis; abundance; medium; energy; strain; recovery; oil
15	Stem cell	differentiation; vitro; stem; culture; proliferation; methylation; adult; lineage; regeneration; capacity; progenitor; stem cell; rise; pluripotent stem; fate; bone marrow; maintenance; expansion; marker; embryonic stem
16	Cellular heterogeneity analysis	analysis; single cell; heterogeneity; level; single-cell; size; cellular heterogeneity; integration; cell population; bulk; volume; chapter; acquisition; individual cell; glioma; drug discovery; cell analysis; tummy; significance; conjunction
17	Single cell living imaging	addition; potential; membrane; release; stimulation; homeostasis; secretion; situ; change; uptake; transport; cytoplasm; fluorescence microscopy; gamma; iuss; fusion; plasma membrane; phosphorylation; cell biology; real time
18	Biofilm formation	formation; structure; environment; behavior; light; body; community; form; plasticity; processing; degree; matrix; shape; nature; length; organization; adaptation; space; fixation; assembly

3.3 Topic evolution analysis based on the LDA model

Topic evolution trends were obtained by calculating topic strengths of each year and the results are listed in Figure 4. The x-coordinate represents the year, and the y-coordinate represents the topic strength. Strengths of some topics are on the rise. Topic3 “Immune response”, topic5 “Phylogeny on single cell level”, topic8 “Molecular mechanism of embryonic development”, topic12 “Cell-to-cell variability analysis” and topic16 “Cellular heterogeneity analysis” show increase tendency in fluctuation. The topic strengths of topic13 “Cancer diagnosis and treatment” is nearly continuously increasing, except for a brief dip in 2015. The strength of some topics are in general decline. Topic4 “Signal transduction”, topic6 “Intracellular calcium modulation”, topic17 “Single cell living imaging” show trends of fluctuating

downward. The topic strength of topic7 “*Single cell gel electrophoresis*” continues to drop. The topic strength of topic0 “*Pathology of brain disease*”, topic1 “*Mathematical modeling of cell cycle*”, topic2 “*Single cell detection platform*”, topic9 “*Cell adhesion*”, topic10 “*Isolation and sorting of single cell*”, topic11 “*Cell migration*”, topic14 “*Single cell oil*”, topic15 “*Stem cell*” and topic18 “*Biofilm formation*” show large volatilities. The strength of almost all the topics refers to the method category tends to decrease since 2015, while the topics refers to the clinical application of single cell technologies such as topic13, 15, and 16 show rapid growth since 2015. This indicates that technological innovation drives the development of applied research.

3.4 Topic distribution analysis of countries

Research topic distribution of countries is usually concerned about in the scientific and technology information analysis. Traditional bibliometric or co-occurrence methods are either too tedious or unable to perform quantitative analysis. Taking the post-discretized method which is used for temporal distribution analysis for reference, the topics can also be dispersed to different countries to detect the spatial distribution. Thus, the research topic distribution can be easily observed through the topic strength analysis. As is shown in Figure 2a, the top10 most productive countries of single cell research field are the US, China, Germany, UK, Japan, France, Canada, Switzerland, Italy, and the Netherlands (from high to low). The topic identification results were dispersed into these ten countries and the topic strengths of each country were calculated. Comparing the topic strengths of every topic in each country, the information about the research investment emphasis and development priorities can be obtained (Figure 5).

The strengths of research topics reflects the country’s focus on different topics or research and development investments. For example, China and Japan paid far more attention to topic2 “*Single cell detection platform*” than other topics. As for Italy, the topic that attracted the most attention was topic7 “*Single cell gel electrophoresis*”. The topic distribution of Germany and the Netherlands are relatively balanced.

3.5 Topic distribution trends of countries

From the two dimensions of time and space, the topic distribution trends of countries can be analyzed. Taking China for example, the topic distribution trend from 2009 to 2019 can be observed in Figure 6. For the convenience of comparison, the topic strength range for each year as 0.04–0.07, with an interval of 0.005. From 2009 to 2013, topic7 “*single cell gel electrophoresis*” was the topic with the highest strength. Since 2014, the superiority of topic7 had been replaced by topic2 “*single cell detection platform*”, and the standard deviation of topic strengths is significantly lower than which in previous years.

Figure 4. Topic strength of 19 topics on single cell research. X-coordinate: Years; Y-coordinate: Topic Strength.

Figure 5. Topic distributions of the top10 most productive countries.

Figure 6. The topic distribution trend of China.

3.6 Topic evolution trends analysis of countries

In terms of a specific topic, the evolution trends of each country can be analyzed. Taking topic13 “*Cancer diagnosis and treatment*” as an example, the evolution trends of each country can be observed from Figure 7. For the convenience of comparison, the topic strength range for each year as 0.042–0.056, with an interval of 0.002. For China, France, and Switzerland, the topic strength increase dramatically since 2015, for Germany, UK, Japan, Canada, and Italy, the topic strength are fluctuate during the last decade.

4 Discussion & conclusion

In order to fully understand the historical progress and current situation, as well as its future development trend of single cell research, this paper conducts a comprehensive bibliometric study based on the publications from WoS between 2009 and 2019. The rapid growth of scientific literature reveals the vigorous development of single cell research in recent years. The top10 most productive countries of single cell research field are the US, China, Germany, UK, Japan, France, Canada, Switzerland, Italy, and the Netherlands. The US takes the leading position in terms of the total publications and total citations in single cell research field.

Topic identification was performed with LDA model and the results were listed in Table 1. The identified topics can be divided into three categories, which include single cell research methods, the mechanisms of biological processes on single cell level, and clinical application of single cell technologies. The topics’ evolution trends were analyzed through post-discretized method by calculating topic strengths in each year. From Figure 4, we propose that “*Immune response*”, “*Phylogeny on single cell level*”, “*Molecular mechanism of embryonic development*”, “*Cell-to-cell variability analysis*”, “*Cancer diagnosis and treatment*”, “*Cellular heterogeneity analysis*” are the most concerned topics with increasing topic strengths. Taking the post-discretized method for reference, the topics were dispersed to the top10 most productive countries and the topic distribution of each country was obtained. For some countries, the topic distributions are relatively balanced, while for the some other countries, the strength of different topics varies greatly. Taking China as an example, this paper also analyzes the trend of a country’s topic distribution changing over time (Figure 6). Before 2013, topic7 “*Single cell gel electrophoresis*” was the topic with the highest topic strength. From 2014 to 2019, topic7 gradually lost its superiority and ranked around the tenth, at the same time topic2 “*single cell detection platform*” became the most focused topic. The evolution trends of each country for a specific topic can also analyzed, this paper takes topic13 “*Cancer diagnosis and treatment*” as an example to illustrate.

Figure 7. Topic evolution trends of the top10 productive countries of topic13.

This paper provides a relatively broad perspective for the evolution of single cell research, and reveals the development trend and hot spots in this field. The topic distribution trends of countries were also analyzed. On the one hand, it can help researchers to grasp the research trend accurately and seize the opportunity of scientific research. On the other hand, it can provide support for national and scientific research institutions to formulate scientific and technological policies and strategic plans.

Acknowledgements

This study was supported by the Chinese Academy of Sciences literature information capability construction project of 2020 “Construction of strategic information research and consultation system in science and technology field” (Grant No. E290001).

Author contributions

Tian Jiang (jiangtian@mail.las.ac.cn): Conceptualization (Lead), Data curation (Lead), Formal analysis (Equal), Investigation (Lead), Methodology (Lead), Validation (Equal), Writing-original draft (Lead), Writing-review & editing (Equal); Xiaoping Liu (liuxp@mail.las.ac.cn): Formal analysis (Equal), Funding acquisition (Lead), Project administration (Supporting), Supervision (Supporting), Validation (Equal), Writing-review & editing (Equal); Chao Zhang (zhangch@mail.las.ac.cn): Formal analysis (Equal), Project administration (Supporting), Resources (Lead), Supervision (Supporting), Validation (Equal), Writing-review & editing (Equal); Chuanhao Yin (cieyinchuanhao@163.com): Data curation (Supporting), Methodology (Equal), Software (Lead); Huizhou Liu (liuhz@mail.las.ac.cn): Formal analysis (Equal), Project administration (Lead), Supervision (Lead), Validation (Equal), Writing-review & editing (Equal).

Reference

- Ai, S.S., Xiong, H.Q., Li, C.C., Luo, Y.J., Shi, Q., Liu, Y.X., et al. (2019). Profiling chromatin states using single cell itChIP-seq. *Nature Cell Biology*, 21(9), 1164–1172.
- Blei, D.M., Ng, A.Y., & Jordan, M.I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3(4/5), 993–1022.
- Briggs, J.A., Weinreb, C., Wagner, D.E., Megason, S., Peshkin, L., Kirschner, M.W., et al. (2018). The dynamics of gene expression in vertebrate embryogenesis at single cell resolution. *Science*, 360(6392). pii: eaar5780.
- Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D.M., Hill, A.J., et al. (2019). The single cell transcriptional landscape of mammalian organogenesis. *Nature*, 566(7745), 496–502.
- Farrell, J.A., Wang, Y., Riesenfeld, S.J., Shekhar, K., Regev, A., & Schier, A.F. (2018). Single cell reconstruction of developmental trajectories during zebrafish embryogenesis. *Science*, 360(6392). pii: eaar3131.
- Griffiths, J.A., Scialdone, A., & Marioni, J.C. (2018). Using single cell genomics to understand developmental processes and cell fate decisions. *Molecular System Biology*, 14(4), e8046.

Research Paper

- Guo, X.Y., Zhang, Y.Y., Zheng, L.T., Zheng, C.H., Song, J.T., Zhang, Q.M., et al. (2018). Global characterization of T cells in non-small-cell lung cancer by single cell sequencing. *Nature Medicine*, 24(7), 978–985.
- Jelodar, H., Wang, Y.L., Yuan, C., Feng, X., Jiang, X.H., Li, Y.C., et al. (2019). Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. *Multimedia Tools and Applications*, 78(11), 15169–15211.
- Junker, J.P., & van Oudenaarden A. (2014). Every cell is special: Genome-wide studies add a new dimension to single cell biology. *Cell*, 157(1), 8–11.
- Lee, L. (2001). On the Effectiveness of the skew divergence for statistical language analysis [C]// Richardson T.S, Jaakkola T.S. In the Proceedings of the English International Workshop on Artificial Intelligence and Statistics. Key West: Society for Artificial Intelligence and Statistics. 65–72.
- Lindström, S. (2012). Flow Cytometry and Microscopy as Means of Studying Single Cells: A Short Introductory Overview. In: Lindström, S., & Andersson-Svahn, H. (Ed.), *Single cell Analysis. Methods in Molecular Biology (Methods and Protocols)*, 853 (pp.13–15). Humana Press.
- Nicolaisen, J. (2010). Bibliometrics and citation analysis: From the science citation index to cybermetrics. *Journal of the American Society for Information Science and Technology*, 61(1), 205–207.
- Reece, A., Xia, B.Z., Jiang, Z.L., Noren, B., McBride, R., & Oakey, J. (2016). Microfluidic techniques for high throughput single cell analysis. *Current Opinion in Biotechnology*, 40, 90–96.
- Wang, Q.H., Xiong H.Q., Ai, S.S., Yu, X.H., Liu, Y.X., Zhang, J.J., et al. (2019) CoBATCH for High-Throughput Single cell Epigenomic Profiling. *Molecular Cell*, 76(1), 206–216.e7.
- Wei, W., Shin, Y.S., Xue, M., Matsutani, T., Masui, K., Yang, H., et al. (2016). Single cell Phosphoproteomics Resolves Adaptive Signaling Dynamics and Informs Targeted Combination Therapy in Glioblastoma. *Cancer Cell*, 29(4), 563–573.
- Zhang, L., Yu, X., Zheng, L.T., Zhang, Y.Y., Li, Y.S., Fang, Q., et al. (2018). Lineage tracking reveals dynamic relationships of T cells in colorectal cancer. *Nature*, 564, 268–272.
- Zhang, W., Xu, X.L., Ming, C.H., Mao, Z.J., Shi, J., & Xiang, Y.Q. (2016). Surviving in the dispute: A bibliometric analysis of global GMF-related research, 1995–2014. *Scientometrics*, 109(1), 359–375.
- Zheng, T.L., Wang, J., Wang, Q.H., Nie, C.H., Shi, Z.N., Wang, X.N., et al. (2016). A bibliometric analysis of micro/nano-bubble related research: Current trends, present application, and future prospects. *Scientometrics*, 109(1), 53–71.

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