Trends of triple negative breast cancer research (2007–2015)
A bibliometric study

Yiran Wang, MDa,b, Xiao Zhai, MDa,c, Chuan Liu, MDd, Ning Wang, MDb,1, Yajie Wang, MDa,∗

Abstract
Background: Triple negative breast cancer (TNBC) is an aggressive breast cancer subtype. However, there have been limited data to evaluate the trend of TNBC research. This study aims to investigate the trend of TNBC research and compare the contribution of research from different regions, organizations, and authors.

Methods: TNBC-related publications from 2007 to 2015 were retrieved from the Web of Science database. Excel 2013 (Redmond, Washington, USA), GraphPad Prism 5 (GraphPad Prism Software Inc., San Diego, CA), and VOSviewer (Leiden University, Leiden, Netherlands) software were used to analyze the trend of TNBC research. This article does not contain any studies with human participants or animals performed by any of the authors.

Results: A total of 1695 papers were identified and were cited 34,078 times with a time limit of May 27, 2016. The United States accounted for 43.10% of the articles, 57.59% of the citations, and the highest H-index (64). China ranked second in total number of articles, but seventh in citation frequency (1998) and ninth in H-index (21). The journal Breast Cancer Research and Treatment had the highest number of publications. The author, Narod SA, has published the most papers in this field (30). The keyword “receptor” was mentioned the most, 1,489 times, and the word “myeloid cell leukemia-1 (MCL-1)” was the latest hot spot by 2015.

Conclusion: Literature growth related to TNBC is expanding rapidly in recent years. The quality of the articles from China still requires improvement. Newest progress of the TNBC research may be released by the journal Breast Cancer Research and Treatment first. Narod SA, Gonzalez-Angulo AM, and Hortobagyi GN may be good candidates for collaborative research in this field. MCL-1 is an emerging topic that should be closely observed.

Abbreviations: BRCA1 = breast cancer 1, ER = estrogen receptor, HER-2 = human epidermal growth factor receptor-2, MCL-1 = myeloid cell leukemia-1, PgR = progesterone receptor, TNBC = triple negative breast cancer.

Keywords: bibliometric, citation, H-index, triple negative breast cancer, VOSviewer

1. Introduction

Triple-negative breast cancer (TNBC) is an aggressive breast cancer subtype and is characterized by negative for estrogen receptors (ERs), progesterone receptors (PgRs), and the human epidermal growth factor receptor-2 (HER-2) gene. It represents about 15% of all cases of breast cancer and occurs more frequently in young and African-American women. Women with TNBC have a poor prognosis due to the aggressive nature and lack of suitable targeted therapies. Presently, research progression for TNBC has not yet been studied. Thus, it is necessary to monitor the global TNBC research.

Bibliometrics is a good choice to evaluate the trend in research activity over time. Bibliometrics uses the literature system and literature metrology characteristics as research objects and analyzes the literatures quantitatively and qualitatively. It can provide an access to characterize the development in a certain field and has played an important role in the past to govern policy making, clinical guideline and research trend in diabetes, cardiovascular disease, respiratory medicine, gastrointestinal diseases. This paper presents a novel study employing the bibliometrics method to analyze TNBC publications retrieved on the Web of Science (Thomson Reuters Company) database from 2007 to 2015. Results were analyzed to better understand the trend of TNBC research and to explore the next possible hot spot in this field.
2. Materials and methods

2.1. Sources of the data and search strategy

A literature search was performed in the Science Citation Index-Expanded (SCI-E) of the Thomson Reuters Web of Science from 2007 to 2015. Based upon the finding that the first paper on TNBC was published in 2007, we restricted our years to 2007 to 2015. The data were downloaded from the public databases, and there were no ethical questions about the data. Ethical approval was not necessary.

All electronic searches were conducted on a single day, May 27, 2016, to avoid changes in citation rate as much as possible. The following search keywords were used: TI = (Triple Negative Breast Cancer) OR TI = (Breast Cancer, Triple-Negative) OR TI = (Breast Neoplasm, Triple-Negative) OR TI = (Triple-Negative Breast Neoplasm) OR TI = (ER Negative PR Negative HER2 Negative Breast Cancer) OR TI = (ER-Negative PR-Negative HER2-Negative Breast Neoplasm) OR TI = (TNBC) OR (TI = (BRCA1) AND TI = (breast) AND TI = (mutation)) AND Language = English. While there are a variety of manuscript types, only peer-reviewed articles, and reviews are included.

2.2. Data collection

The data were extracted carefully from all eligible publications independently by 2 authors (YW and XZ). The text data downloaded from Web of Science were imported into Microsoft Excel 2013, GraphPad Prism 5 (GraphPad Prism Software Inc., San Diego, CA), and VOSviewer (Leiden University, Leiden, The Netherlands). The data were analyzed both quantitatively and qualitatively.

2.3. Statistical methods

Web of Science was used to analyze the characteristics of the publications, including countries and regions, time, authors, organizations, citation frequency, and H-index. The relative research interest was calculated as weighted publications per year divided by all weighted publications per year.

GraphPad Prism 5 was used to analyze the time trend of the publications. The logistic regression model, \( f(x) = \frac{c}{1 + ae^{-bx}} \), was used to calculate the cumulative volume and to predict future trend of papers in this field. The inflection point of the logistic curve was the time when the growth rate of papers moved from positive to negative. The formula \( T = \ln \frac{a}{b} \) was used to generate the point.

VOSviewer was used to analyze the relations among highly cited references and productive authors. It is commonly used for co-citation network analysis and visualization.\(^{11}\) The VOSviewer was also used to generate the knowledge maps of cited references and keywords related to TNBC research.

3. Results

3.1. Countries/regions contributing to global publications

A total of 1695 articles met the search criteria from 2007 to 2015. The United States published the most papers (731, 43.13%), followed by China (258, 15.22%), and Italy (123, 7.26%) (Fig. 1A). The United States published the most papers per year (Fig. 1B). The institute with the greatest number of publications was University of Texas M.D. Anderson Cancer Center with a total of 87 papers, accounting for 5.13% of all published literature relating to the field. There were 13 American institutes in the top 20 institutes list. Meanwhile, 3 Chinese institutes, 2 Canadian institutes, 1 English institute, and 1 Korean institute were on the list. Publications from top 20 institutes accounted for 42.07% of all literature on TNBC (Fig. 3B).

3.2. Citation and H-index analysis

According to the analysis of the Web of Science database, all articles related to TNBC had been cited 34,078 times since 2007 (24,565 times without self-citations). The cited frequency per paper was 20.11 times. The number of citations of papers from the United States was 19,624, accounting for 57.59% of the total citations. The H-index of papers from the United States was 64. England ranked second with the citation frequency of 5645 and the H-index of 34. Though the number of publications of China ranked second, the citation frequency and H-index ranked the seventh and ninth, respectively (Fig. 3A).

3.3. Distribution of institutes paying attention to TNBC

The institute with the greatest number of publications was University of Texas M.D. Anderson Cancer Center with a total of 87 papers, accounting for 5.13% of all published literature relating to the field. There were 13 American institutes in the top 20 institutes list. Meanwhile, 3 Chinese institutes, 2 Canadian institutes, 1 English institute, and 1 Korean institute were on the list. Publications from top 20 institutes accounted for 42.07% of all literature on TNBC (Fig. 3B).

3.4. Distribution of published journals on TNBC

Nearly half of the publications were published in the top 20 journals (769, 45.37%). The journal Breast Cancer Research and Treatment (IF = 3.94, 2016) published the most with 142 papers.
There was a case report, an article, and a review in the *New England Journal of Medicine* (IF = 55.876, 2016) on TNBC. Two reviews and 7 articles were in the *Lancet Oncology* (IF = 45.217, 2016). The top 20 journals that published the most papers are shown in Fig. 3C.

### 3.5. Distribution of authors on TNBC

Top 10 authors contributed a total of 217 papers relating to TNBC, accounting for 12.8% of all published literature relating to the field. Narod SA published the most papers in this field (30 papers), followed by Gonzalez Angulo AM with 28 publications and Hortobagyi GN with 24 publications (Table 1).

### 3.6. References analysis on TNBC

The reference analysis is one of the most important indicators of bibliometrics. References of papers in the database were analyzed by VOSviewer, and the top 74 references which were cited more than 50 times in the database were selected for analysis. The references included in the analysis were divided into 3 clusters (Fig. 4A).

The first cluster included 30 papers and mainly focused on the clinical features of TNBC. The second cluster included 25 papers and focused on molecular identification of TNBC. The third cluster included 19 papers and was related to the treatment of TNBC (Fig. 4) (Supplemental table 1, http://links.lww.com/MD/B412).

### 3.7. Hot spots of studies on TNBC

Keywords used in the 1695 papers found during the methodology search were analyzed through VOSviewer. As shown in Fig. 5, the 400 keywords (defined as being used more than 20 times within titles and abstracts in all of articles) were classified into 4 clusters: “molecular mechanism,” “clinical research,” “treatment,” and “risk factor” (Fig. 5A). Among the “molecular mechanism” cluster, keywords used in the publications of TNBC were listed as follows: cell (1254 times), effect (721 times), MDA-MB (535 times).
times), metastasis (507 times), and pathway (469 times). For the cluster of clinical researches, the primary keywords were as follows: receptor (1489 times), year (541 times), rate (418 times), overall survival (342 times), and stage (309 times). For the cluster of treatment, the main keywords were as follows: month (363 times), combination (359 times), trial (280 times), grade (266 times), and efficacy (238 times). For the cluster of risk factor, the keywords were as follows: women (864 times), breast cancer 1 (BRCA1) (772 times), mutation (587 times), risk (513 times), and age (331 times). The results demonstrated that the dominant fields of TNBC include 4 directions (Supplemental table 2, http://links.lww.com/MD/B412).

In Fig. 5B, VOSViewer applied colors to keywords based upon when they appeared in literature. Keywords red in color appeared early, and keywords yellow in color appeared later. In early stage of TNBC research, clinical feature of the TNBC was the main hot

| Author       | Number of papers | Country | Affiliations                                                                 |
|--------------|------------------|---------|------------------------------------------------------------------------------|
| Narod SA     | 30               | Canada  | Women’s College Research Institute, University of Toronto                     |
| Gonzalez-Angulo AM | 28        | United States | Department of Breast Medical Oncology, The University of Texas M.D. Anderson Cancer Center |
| Hortobagyi GN | 24               | United States | Department of Breast Medical Oncology, The University of Texas M.D. Anderson Cancer Center |
| Lubinski J   | 23               | Poland  | Hereditary Cancer Center, Pomeranian Medical University                       |
| Shao ZM      | 22               | China   | Department of Breast Surgery, Breast Cancer Institute, Fudan University        |
| Sun P        | 20               | Canada  | Women’s College Research Institute, Women’s College Hospital and University of Toronto |
| Carey LA     | 18               | United States | Hematology/Oncology, University of North Carolina–Lineberger Comprehensive Cancer Center, |
| Faulkes WD   | 18               | Canada  | Program in Cancer Genetics, Department of Oncology, McGill University         |
| Lynch HT     | 18               | United States | Department of Preventive Medicine and Public Health, Creighton University School of Medicine, Omaha, Nebraska |
| Li J         | 16               | China   | Key Laboratory of Carcinogenesis and Translational Research, Breast Center, Peking University Cancer Hospital & Institute |

Table 1
Top 10 authors with the most publications related to TNBC research.

Figure 4. The reference analysis of the top 74 papers. (A) Mapping on co-cited references of TNBC. Due to the large number of cited references, only papers that were cited more than 50 times (N = 74) were included in the analysis. (Note: the line between every 2 points within the figure represents that both papers were cited in 1 article. If the line is thicker, the link between 2 papers is closer.). (B) Mapping on density visualization of co-cited references. Different colors represent different co-cited times of papers. The color of an item was determined by the co-cited times, where by default colors range from blue (few times) to green (average times) to red (many times). Using this system, highly cited papers were shown in red within the density visualization.
spot. Recent trend showed words of “myeloid cell leukemia-1” (MCL-1) and “nanoparticle” appeared in 2014 as keywords for 23 and 58 times, respectively. These words both belonged to the first (“molecular mechanism”) cluster. In the third (“treatment”) cluster, “mTNBC” and “circulating tumor cells” appeared in 2014. The number of occurrences of them was 31 and 59, respectively. There was no new word in the second (“clinical researches”) cluster and fourth (“risk factor”) cluster appearing in 2014 (Supplemental table 2, http://links.lww.com/MD/B412). Meanwhile, the keyword “receptor” is in center of Fig. 5B, and it co-occurred with other words for 30,487 times, followed by woman with 22,424 times of co-occurrences with.

4. Discussion

4.1. Trend of TNBC research

The United States and England ranked first and second for citation frequency in TNBC research respectively, and the number of publications from the United States was higher than that of any other countries. One of the reasons might be that there are many African-American women patients in the United States, and American government have invested much money on this disease. In addition, a number of citations and H-index of the publications from the United States were higher than that of any other countries or regions, suggesting that American scientists have taken the leading position regarding TNBC research. However, China ranked second in total number of articles for many years, but seventh in citation frequency and ninth in H-index. Furthermore, it showed that improvements are necessary to keep up with publications in the field.

In addition, there was a rapid growth of publications related to TNBC research from 2007 to 2012 all over the world. The inflection point of China was later than that of other countries because its research started later than that of the United States, India, and other countries. When the growth of the publications gets the inflection point, the studies in this field are relatively deep and mature. There still will be new publications in this field every year, but the absolute number is decreasing year by year. This finding showed that new development in TNBC field would be more difficult than before.

Meanwhile, Italy, Canada, England, and Germany had fewer publications than that of China, but their citation frequency and H-index were higher than that of China. This suggested that more attention needed to be taken to improve the quality of Chinese research once again.

American institutes were the leading organizations on TNBC research. More than half of the top 20 institutes were in the United States. Meanwhile, 3 institutes were in China. The powerful institutions were several American elite institutions. The reason why the United States had the most publications in this field was that it had the most powerful institutions.

It is of note that the Journals Breast Cancer Research and Treatment, PLoS ONE, Oncotarget, and Breast Cancer Research...
were the main journals involved in publishing TNBC papers. Furthermore, it indicated that future development within TNBC would likely be showcased within the aforementioned journals.

In terms of authors and publications, Narod SA, Gonzalez-Angulo AM, and Hortobagyi GN had published the most articles on TNBC. Narod et al.\(^2\) focused on the clinical study of BRCA1 mutation carriers. Gonzalez-Angulo et al.\(^3\) focused on clinical research on TNBC. Hortobagyi et al.\(^4\) emphasized the molecular mechanism and treatment of TNBC. These scientists were leaders in the TNBC research, and their studies may still have a huge impact on TNBC research in future and help design our own experiments.

4.2. Researches focused on TNBC

Articles that were cited the most on TNBC were the fundamental basis for further studies. In the cluster of “molecular mechanism,” the article titled “Basal-like breast cancer defined by five bio-markers has superior prognostic value than triple-negative phenotype” was cited for 556 times. It was published in Clinical Cancer Research in 2008. The article found that the expanded surrogate immunopanel of ER, PgR, HER-2, epidermal growth factor receptor, and cytokeratin 5/6 provided a more specific definition of basal-like breast cancer that better predicted breast cancer survival.\(^5\) In the cluster of “treatment,” the article “The triple negative paradox: primary tumor chemosensitivity of breast cancer subtypes” was cited for 796 times which was published in Clinical Cancer Research too. It concluded that basal-like and HER2+/ER− subtypes were more sensitive to anthracycline-based neoadjuvant chemotherapy than luminal breast cancers. Patients that had pathologic complete response to chemotherapy had good prognoses regardless of subtype.\(^6\) In the cluster of “clinical features,” the article “Triple-negative breast cancer: clinical features and patterns of recurrence” published in Clinical Cancer Research back in 2007 was subsequently cited 1041 times. The review compared the clinical features, natural history, and outcomes for women with “triple-negative” breast cancer with women with other types of breast cancer and concluded that TNBC had a more aggressive clinical course than other forms of breast cancer, but the adverse effect was transient.\(^7\)

Most publications were related to the molecular mechanism and treatment of TNBC while a majority of researches also emphasized the risk factors. Their studies solved the biggest, most difficult, and most pressing problems at that time. At the same time, the new classification of TNBC that better predicts breast cancer prognosis is widely applied to clinical practice. Anthracycline-based neoadjuvant chemotherapy can be used to improve treatment efficacy for HER2+/ER− patients. In addition, some studies showed that TNBC had an aggressive course and revealed the differences between TNBC and other types of breast cancer. These studies were the milestones. The results played an important role in promoting progress in TNBC research. Moreover, the bibliometric analysis showed that MCL-1 was the next hot spot in this field. Scientists in this field may acquire some research achievements and publish high-level papers.

4.3. Strengths and limitations

Papers on TNBC evaluated in this study were reviewed from the Web of Science database of SCI-E journals. The data analysis was relatively comprehensive and objective. However, there are some limitations. Papers in non-English languages may be not included in the database and analyze and may have excluded important non-English research studies in TNBC. However, the Web of Science is the largest bibliometric database currently available. In addition, papers published in 2016 were not included in this article, so the analysis of hot spots did not cover the keywords of 2016. Future work should address studies in other non-English languages. The publications in 2016 can be analyzed in 2017.

5. Conclusion

In conclusion, this study helps scientists master the trends of TNBC research. There was a mismatch between the influence and the quantity of publications from China. Although China has published a considerable number of articles on the topic, improvements are necessary to keep up with publications in the field. Scientific research in TNBC field would be fewer and more difficult than before. Newest progress of the TNBC research could be found in Breast Cancer Research and Treatment. Narod SA, Gonzalez-Angulo AM, and Hortobagyi GN may be good candidates for collaborative research in this field. MCL-1 may be the latest hot spot in the TNBC research and individuals currently involved in it may be pioneers to lead the field in TNBC research in the next few years.

References

1. Dietze EC, Sistrunk C, Miranda-Carboni G, et al. Triple-negative breast cancer in African-American women: disparities versus biology. Nat Rev Cancer 2015;15:248–54.
2. Sturtz LA, Melley J, Mammul K, et al. Outcome disparities in African American women with triple negative breast cancer: a comparison of epidemiological and molecular factors between African American and Caucasian women with triple negative breast cancer. BMC Cancer 2014;14:62.
3. Cabezón T, Gromova I, Gromov P, et al. Proteomic profiling of triple-negative breast carcinomas in combination with a three-tier orthogonal technology approach identifies mage-A4 as potential therapeutic target in estrogen receptor negative breast cancer. Mol Cell Proteomics 2013;12:381–94.
4. Ekinci S, Agili M, Ersen O, et al. Letter to the editor regarding analysis of changing paradigms of management in 179 patients with spinal tuberculosis during a 12-year period and proposal of a new management algorithm. World Neurosurg 2015;84:2072.
5. Avcu G, Bal ZS, Duyu M, et al. Thanks to trauma: a delayed diagnosis of Pott disease. Pediatr Emerg Care 2013;31:E17–8.
6. Geaney F, Scaturro G, Kelly C, et al. Type 2 diabetes research yield, 1951–2012: bibliometrics analysis and density-equalizing mapping. PLoS One 2015;10:e0133009.
7. Shuaib W, Khan MS, Shahid H, et al. Bibliometric analysis of the top 100 cited cardiovascular articles. Am J Cardiol 2015;115:972–81.
8. Seriwala HM, Khan MS, Shuaib W, et al. Bibliometric analysis of the top 50 cited respiratory articles. Expert Rev Respir Med 2015;9:817–24.
9. Narotsky D, Green PH, Lembold B. Temporal and geographic trends in celiac disease publications: a bibliometric analysis. Eur J Gastroenterol Hepatol 2013;25:1071–7.
10. Hagen AJ, Bofin AM, Ytterhus B, et al. Amplification of TOP2A and HER-2 genes in breast cancers occurring in patients harbouring BRCA1 germline mutations. Acta Oncol 2007;46:199–203.
11. Symnevedt MB, Chen C, Holmes JH. CiteSpace II: visualization and knowledge discovery in bibliographic databases. AMIA Annu Symp Proc 2005;2005:724–8.
12. Gromwald J, Robidoux A, Kim-Sing C, et al. Duration of tamoxifen use and the risk of contralateral breast cancer in BRCA1 and BRCA2 mutation carriers. Breast Cancer Res Treat 2014;146:421–7.
13. Narod SA, Metcalfe K, Lynch HT, et al. Should all BRCA1 mutation carriers with stage I breast cancer receive chemotherapy? Breast Cancer Res Treat 2015;153:273–9.
14. Gonzalez-Angulo AM, Timms KM, Liu SY, et al. Incidence and outcome of BRCA mutations in unselected patients with triple receptor-negative breast cancer. Clin Cancer Res 2011;17:1082–9.
15. Dawood S, Lee XD, Litton JK, et al. Impact of body mass index on survival outcome among women with early stage triple-negative breast cancer. Clin Breast Cancer 2012;12:364–72.
[16] Bartholomeusz C, Gonzalez-Angulo AM, Liu P, et al. High ERK protein expression levels correlate with shorter survival in triple-negative breast cancer patients. Oncologist 2012;17:766–74.

[17] Gonzalez-Angulo AM, Akcakanat A, Liu S, et al. Open-label randomized clinical trial of standard neoadjuvant chemotherapy with paclitaxel followed by FEC versus the combination of paclitaxel and everolimus followed by FEC in women with triple receptor-negative breast cancer. Ann Oncol 2014;25:1122–7.

[18] Cheang MCU, Voduc D, Bajdik C, et al. Basal-like breast cancer defined by five biomarkers has superior prognostic value than triple-negative phenotype. Clin Cancer Res 2008;14:1368–76.

[19] Carey LA, Dees EC, Sawyer L, et al. The triple negative paradox: primary tumor chemosensitivity of breast cancer subtypes. Clin Cancer Res 2007;13:2329–34.

[20] Dent R, Trudeau M, Pritchard KI, et al. Triple-negative breast cancer: clinical features and patterns of recurrence. Clin Cancer Res 2007;13:4429–34.