Postprint: Bibliometric Analysis of RNA Sequencing Technology in Liver Cancer

Authors: Ge Qiong, Hu Jiakang, Yu Yuqi, Lai Wenwen, Luo Shiwen, Lu Quqin, Lu Quqin

Date: 2023-10-24T00:00:00+00:00

Abstract

Background In recent years, the extensive application of transcriptome sequencing technology in liver cancer research has facilitated genomic and biological studies of this malignancy. A comprehensive summary and analysis of literature on RNA-seq applications in liver cancer research over the past two decades can help researchers gain a thorough understanding of research hotspots and recent advances in this field, thereby providing valuable references for future studies.

Objective To conduct a holistic evaluation of RNA-seq applications across various aspects of liver cancer research, including therapy, diagnosis, and pathogenesis, through bibliometric analysis; to reveal the global distribution of research hotspots in the field of RNA-seq applied to liver cancer, and to speculate on future development trends.

Methods English-language literature on RNA-seq applications in liver cancer research indexed in the Web of Science database from January 1, 2001 to December 31, 2022 was retrieved. Microsoft Excel 2016 was employed for publication output analysis, while CiteSpace software was utilized for visual analysis of authors, countries, institutions, and keywords. Visualization results were evaluated using metrics including publication output, centrality, modularity value, and average silhouette value.

Results A total of 1,397 publications on RNA-seq applications in liver cancer research were identified in the database. Analysis revealed that a core author group has emerged in this field. China demonstrated the highest publication output, though with slightly less research depth; the United States ranked second in publication volume but exhibited the highest centrality (0.44). Institutions with high publication output were predominantly located in China, whereas those with greater influence were mainly distributed in the United States. Keyword analysis indicated that research hotspots in this field focus on molecular

mechanisms underlying liver cancer development and progression, gene expression, and molecular markers.

Conclusion Molecular mechanisms of liver cancer development and progression, disease biomarkers, and therapeutic targets constitute the current research hotspots in this field, with clinical precision medicine potentially emerging as a key future research direction.

Full Text

Bibliometric Analysis of RNA-seq Technology in Liver Cancer Research

GE Qiong[12], HU Jiakang[12], YU Yuqi[12], LAI Wenwen[12], LUO Shiwen[3], LU Quqin[12]*

[1]School of Public Health, Nanchang University, Nanchang 330006, China
[2]Jiangxi Province Key Laboratory of Preventive Medicine, Nanchang 330006, China
[3]The First Affiliated Hospital of Nanchang University, Nanchang 330006, China

Corresponding author: LU Quqin, Professor, Doctoral Supervisor; E-mail: quqinlu@ncu.edu.cn

Abstract

Background In recent years, the extensive application of transcriptome sequencing technology in liver cancer research has significantly advanced genomic and biological studies of this disease. Summarizing and analyzing the literature on RNA-seq applied to liver cancer research over the past two decades can help researchers gain a comprehensive understanding of research hotspots and latest developments in this field, providing valuable references for future studies.

Objective To conduct a holistic evaluation of RNA-seq applications across various aspects of liver cancer research—including treatment, diagnosis, and pathogenesis—through bibliometric analysis, thereby revealing the global distribution of research hotspots in this domain and predicting future development trends.

Methods We searched the Web of Science database for English-language literature on RNA-seq applied to liver cancer research published between January 1, 2001, and December 31, 2022. Microsoft Excel 2016 was used to analyze publication volumes, while CiteSpace software enabled visualization of authors, countries, institutions, and keywords. Evaluation metrics included publication count, centrality, modularity value, and average silhouette value.

Results A total of 1,397 relevant documents were retrieved. The analysis revealed that a core author group has emerged in this field. China contributed the highest number of publications, though with slightly less research depth,

while the United States ranked second in publication volume but demonstrated the highest centrality at 0.44. Most high-output institutions were located in China, whereas the most influential institutions were predominantly in the United States. Keyword analysis indicated that research hotspots focus on molecular mechanisms of liver cancer development, gene expression, and molecular markers.

Conclusion Current research hotspots in this field include the molecular mechanisms of liver cancer pathogenesis, disease biomarkers, and therapeutic targets. Clinical precision medicine is likely to become a key research direction in the future.

Keywords Liver cancer; RNA-seq; Research hotspots; Research trends; CiteSpace; Bibliometric analysis; Visualization

---

Introduction

Primary liver cancer is one of the most common malignant tumors worldwide. According to statistics, there were 905,677 new cases of primary liver cancer globally in 2020, accounting for 4.7% of all new cancer cases and ranking sixth among all cancers; deaths reached 830,180, representing 8.3% of all cancer-related deaths and ranking third among cancer mortality causes [1]. Over recent decades, researchers have made significant progress in liver cancer diagnosis, treatment, and prognosis, particularly with the application of gene sequencing technologies such as RNA-seq, which has led to the continuous discovery of pathogenic mechanisms and diagnostic/prognostic biomarkers, thereby improving outcomes for liver cancer patients to some extent.

RNA sequencing (RNA-seq) is a well-established transcriptomics research method that utilizes high-throughput sequencing to obtain transcriptome information from cellular tissues. It can be applied to study different RNA populations, including messenger RNA and non-coding RNA, facilitating the discovery of novel transcripts, identification of alternatively spliced genes, and detection of allele-specific expression [2]. In recent years, transcriptomics has been employed to identify disease biomarkers and investigate biological responses to various stimuli and stressors, playing a crucial role in genomic and molecular biology research [3].

With the widespread application of transcriptome sequencing technology, numerous high-quality studies on liver cancer RNA-seq have emerged. The information contained in these publications comprehensively reflects the development process and trending topics in this research field. Summarizing previous findings can positively influence future research directions and disciplinary development. This study employs CiteSpace 6.1.R6 software [4] to conduct a comprehensive bibliometric analysis of liver cancer RNA-seq research literature from the Web of Science (WOS) database over the past two decades, aiming to understand

current research hotspots and future development trends to provide references for subsequent studies.

Methods

1.1 Data Sources and Search Strategy Statistical analysis data were sourced from the Web of Science Core Collection, with the citation index set to "Science Citation Index Expanded (SCI-EXPANDED)." To ensure comprehensive literature retrieval, we used a general search with the following search string: (TS="liver cancer" OR "hepatocellular carcinoma" OR "hepatic carcinoma" OR "hepatoma" OR "hepatocarcinoma" OR "HCC" OR "hepatic cancer" OR "liver carcinoma") AND (TS="RNA sequencing" OR "RNA-seq" OR "RNA sequence"). The search was limited to English-language "Article" document types published between January 1, 2001, and December 31, 2022.

To ensure data quality and representativeness, two research team members independently screened the literature based on content relevance to the themes of "liver cancer" and "RNA-seq," excluding unrelated documents. After verification, comparison, and deduplication, 1,397 articles were retained for final analysis. Data were exported in "plain text file" format with "full record and cited references" selected, including title, keywords, authors, institutions, addresses, abstracts, and publication dates for each document.

1.2 Visualization Analysis Methods 1.2.1 Visualization Approach
Using CiteSpace 6.1.R6 software, we converted the retrieved literature on RNA-seq applications in liver cancer research into analyzable data formats. Annual publication volumes were counted and visualized using Excel software to plot trends in yearly publications and cumulative publications. CiteSpace was employed for network visualization analysis of authors, countries, and institutions, as well as for co-occurrence, clustering, and timeline analyses of keywords.

1.2.2 Parameter Settings
The time span was set from January 2001 to December 2022, with a time slice of 2 years. The g-index threshold parameter was set between 5 and 25 based on node characteristics, and TopN% was uniformly set at 10%.

1.2.3 Data Analysis and Evaluation Criteria
Visualization results were evaluated using metrics including publication count, betweenness centrality, modularity value (Q), and average silhouette value (S). Betweenness centrality measures node importance in a network; higher values indicate greater importance [5]. A centrality exceeding 0.1 identifies a central node that is significant and influential in the research field. The modularity value Q and average silhouette value S assess clustering effectiveness: Q measures clustering quality (higher values indicate better partitioning), while S measures cluster member homogeneity (higher values indicate greater similarity among cluster members). Generally, $Q > 0.3$ and $S > 0.7$ indicate well-distributed clustering.

Results

2.1 Publication Trend Analysis of RNA-seq Applications in Liver Cancer Research Publication volume is the most fundamental bibliometric indicator, encompassing annual and cumulative publication counts [6]. Between 2001 and 2022, we retrieved 1,397 publications on RNA-seq applications in liver cancer research. Annual publication statistics revealed two distinct phases: from 2001 to 2016, publications increased gradually from 4 to 48 (steady growth phase); from 2017 to 2022, publications surged from 80 to 366 (rapid growth phase). Exponential fitting of the cumulative publication data yielded a goodness-of-fit $R^2 = 0.977$, indicating exponential growth in cumulative publications from 2001 to 2022 (Figure 1 [Figure 1: see original paper]).

2.2 Author Analysis of RNA-seq Applications in Liver Cancer Research Author collaboration network maps reveal relationships among researchers in the RNA-seq liver cancer field. CiteSpace was used to select authors in the top 10% by publication volume and generate an author co-occurrence network map (Figure 2 [Figure 2: see original paper]). The network comprised 377 nodes and 774 connections, with a density of 0.0109. Nodes represent authors (larger nodes indicate more publications), while connections represent collaborations (thicker lines indicate stronger cooperation). Overall, researchers were relatively dispersed without forming prominent collaborative networks.

Additionally, author publication counts help identify core research groups. According to Price's Law, the formula for core authors is:

$$M \approx 0.749\sqrt{N_{\max}}$$

where M represents the publication threshold and Nmax represents the maximum number of papers by a single author. Authors exceeding M publications are considered core authors; when their total publications reach 50% of all papers in the field, a core author group has formed.

In our dataset, CHEN Gang had the highest publication count (20 papers), giving Nmax = 20. Calculating M yielded 3.35; thus, authors with more than 3 publications (rounded to 3) were identified as core authors. Analysis identified 180 core authors who collectively published 834 papers, representing 69.6% of the total publications in this field, confirming that a core author group has indeed formed.

The top-ranked core author, CHEN Gang (20 publications), leads a team [7] that primarily utilizes sequencing data from liver cancer patients to investigate pathogenic mechanisms, including molecular-level regulatory mechanisms involving microRNA, lncRNA, and certain proteins, and examines the clinical significance of these markers in liver cancer diagnosis and prognosis. The second-ranked scholars, ZHOU Jian and FAN Jia (17 publications each), lead

a team [8] that focuses on using single-cell sequencing data to study signaling pathways related to hepatocellular carcinoma metastasis, recurrence, growth, and invasion, providing insights for clinical treatment.

2.3 Country Analysis of RNA-seq Applications in Liver Cancer Research A country (region) collaboration network map was generated using CiteSpace (Figure 3 [Figure 3: see original paper]), comprising 55 nodes and 242 connections with a density of 0.163. Nodes represent countries/regions, and connections represent collaborative relationships. The relatively high network density indicates complex international collaboration networks, suggesting active cooperative research relationships among nations.

Analysis was conducted from two perspectives: publication volume and betweenness centrality. In terms of publication count, China ranked first with 990 publications (54.94% of total), followed by the United States with 261 publications. Japan, South Korea, and Germany rounded out the top five, each with more than 50 publications. Regarding betweenness centrality, nodes with purple halos in Figure 3 indicate centrality values exceeding 0.1, including the United States (0.44), Germany (0.30), England (0.23), Brazil (0.14), and Italy (0.13).

2.4 Institution Analysis of RNA-seq Applications in Liver Cancer Research Analyzing publication patterns across institutions reveals the distribution of research capacity in RNA-seq applications for liver cancer. An institutional collaboration network map was created using CiteSpace (Figure 4 [Figure 4: see original paper]), containing 315 nodes and 1,263 connections with a density of 0.0255. The top five institutions by publication volume were all Chinese: Fudan University (80 publications), Sun Yat-sen University (78 publications), Naval Medical University (64 publications), and Guangxi Medical University (64 publications). Institutions with centrality exceeding 0.1 included Harvard University (0.15), Naval Medical University of the Chinese People's Liberation Army (0.14), and Icahn School of Medicine at Mount Sinai (0.14).

Geographical distribution of institutions corresponded to national publication patterns: high-output institutions were predominantly in China, while more influential institutions were primarily in the United States.

2.5 Keyword Co-occurrence Analysis of RNA-seq Applications in Liver Cancer Research Keywords provide high-level summaries of article themes, representing core content. Their frequency, relevance, and burstiness can reveal research hotspots, internal connections, and significance within a field. A keyword co-occurrence network map was generated using CiteSpace (Figure 5 [Figure 5: see original paper]), comprising 454 nodes and 828 connections with a density of 0.0081. The top 20 keywords by frequency are listed in Table 3 . Keywords appearing more than 100 times included hepatocellular carcinoma, expression, cancer, cells, proliferation, liver cancer, metastasis, identification, gene expression, growth, gene, activation, and progression. These indicate

that research hotspots primarily encompass molecular mechanisms related to liver cancer development, proliferation, recurrence, and metastasis, along with gene expression and molecular markers.

2.6 Keyword Clustering Analysis of RNA-seq Applications in Liver Cancer Research Keyword clustering maps illustrate different research focuses within the field and reflect the composition of various research themes. CiteSpace's default EM (Expectation Maximization) clustering algorithm was applied to the keyword network, with the LLR algorithm extracting cluster labels from keywords. The resulting clustering map achieved S = 0.8966 and Q = 0.7688, indicating a well-structured network that effectively represents research hotspots in the field. The keyword clustering knowledge map contained 19 clusters (Figure 6 [Figure 6: see original paper]), primarily covering: (1) macroscopic etiological studies of liver cancer, such as hepatitis C virus and non-alcoholic fatty liver disease; (2) microscopic etiological studies, such as gene mutations and altered signaling pathways; (3) pathogenic mechanism studies, including the role of PI3K/Akt signaling pathways in liver cancer development; (4) research on the liver cancer immune microenvironment; and (5) studies on diagnostic and prognostic biomarkers and therapeutic targets.

2.7 Keyword Timeline Analysis of RNA-seq Applications in Liver Cancer Research Timeline maps reveal the developmental trajectory of a research field, helping investigators understand the temporal evolution of important keywords. Using CiteSpace, a timeline map was generated based on keyword clustering, resulting in 178 nodes and 922 connections (Figure 7 [Figure 7: see original paper]). Line colors represent cluster emergence times, nodes on lines represent keywords within clusters, node sizes indicate keyword frequency, and node positions show first-appearance dates. Overall, the most intensive period for keyword emergence was 2015–2022, coinciding with the rapid growth phase in publication volume.

Discussion

The volume of literature related to a specific topic reflects researcher attention—more publications indicate more active research. Our analysis shows that liver cancer RNA-seq research increased steadily from 2001 to 2022, with the cumulative growth curve following an "S" shape. The field is currently in the first half of this developmental trajectory, with steady growth from 2001–2016 and rapid expansion from 2017–2022. According to Price's logistic growth curve and Thomas Kuhn's theory of scientific development patterns [4], this field is in the "paradigm accumulation" stage, transitioning from theoretical to applied research. We can anticipate that future studies will increasingly focus on applications, particularly using RNA-seq research findings for clinical liver cancer treatment.

Analysis of country/region, institution, and author distributions enhances un-

derstanding of team collaboration and global cooperation, helping researchers identify resources to improve collaborative efficiency. The author collaboration network showed low density, with researchers relatively dispersed, suggesting that enhanced cooperation among investigators in this field is needed. From a national perspective, comprehensive analysis of publication volume and centrality reveals that while China leads in quantity (54.94% of total), its centrality is only 0.1, likely due to large researcher numbers but insufficient research depth. The United States ranks second in publication volume but first in centrality (0.44), indicating strong research capacity in RNA-seq applications for liver cancer. Among the top 10 countries by publication volume, five are in Europe, four in Asia, and one in North America, with developed countries comprising 90%. Influential countries (centrality > 0.1) are primarily in North America (United States) and Europe (Germany, United Kingdom, Brazil, and Italy), a pattern that aligns with institutional distributions. This suggests that while researchers in both East and West are engaged in liver cancer transcriptomics, Western scholars emphasize collaboration more and conduct deeper research, possibly because better socioeconomic development ensures adequate funding, resources, and human capital for exploring novel scientific questions [9]. Consequently, biomedical research in Europe and America started earlier with stronger foundations. Although China started later in cancer biology research and thus has less influence, increased national strength and research investment have gradually boosted its publication output. China should strengthen exchanges among nations, institutions, and authors to improve research quality and enhance its impact in this field.

Keyword analysis, combined with detailed literature review, provides insights into developments in both liver cancer research and transcriptome sequencing technology. Our keyword analysis reveals that recent years have witnessed qualitative leaps in exploring liver cancer risk factors, molecular mechanisms of disease occurrence and progression [10-11], diagnostic and prognostic markers [12], and therapeutic targets [13], driven by increasingly widespread RNA-seq application. Regarding risk factors, studies have shown that hepatitis B is a risk factor for liver cancer, with its mutagenic effects resulting from viral DNA integration into human hepatocytes, forming chimeric fusion transcripts that induce carcinogenesis [14]. In terms of biomarkers, discoveries of various genes [15] and circRNAs [16] have enabled more accurate management of HCC patients in clinical trials and practice, while also providing directions for targeted therapy drug development. Liver cancer transcriptome sequencing analysis also facilitates precision medicine [17], with integration of large-scale genomic datasets with functional and other omics data offering new perspectives for treatment. Conversely, applications in liver cancer have also advanced transcriptome sequencing technology. Technologically, single-cell sequencing has emerged as an improvement over traditional bulk sequencing, providing a solid foundation for precise research [18]. Analytically, continuously updated algorithms have optimized research outcomes [19]. Furthermore, research approaches to transcriptome sequencing data have evolved from initial “dry lab” bioinformatic analyses to

validation through immunohistochemistry and QT-PCR experiments, and more recently to using sequencing data to test experimental hypotheses followed by experimental validation of analytical results [20]. This shift in research thinking has enabled RNA-seq methods to deliver different values. We predict that future research will emphasize integrating sequencing data with experimental studies, with clinical applications focusing on achieving precision treatment for liver cancer through RNA-seq.

In the era of big data analytics, literature information across fields is proliferating. Through bibliometric analysis, researchers can intuitively understand hotspot trends and clarify research directions [21]. Based on CiteSpace visualization software, our analysis of WOS database literature from 2001–2022 on liver cancer RNA-seq research reveals exponential growth in global publications, with molecular mechanisms of liver cancer development, disease biomarkers, and therapeutic targets representing current hotspots. However, Chinese institutions conduct relatively dispersed research with slightly insufficient depth, suggesting that future efforts should strengthen international collaboration to improve research quality.

This study has limitations: it only included English-language literature from the WOS database, excluding high-quality studies in other languages and databases, which may introduce bias. Future research should consider incorporating relevant literature from CNKI, Wanfang, PubMed, and other databases to obtain more comprehensive results.

Author Contributions

GE Qiong was responsible for study design, data analysis and interpretation, and manuscript writing and revision. HU Jiakang and LAI Wenwen collected, screened, and verified literature data. YU Yuqi performed statistical analysis and created figures and tables. LU Quqin and LUO Shiwen provided funding support, revised the manuscript, and ensured quality control, taking overall responsibility for the article.

This article has no conflicts of interest.

References

[1] SUNG H, FERLAY J, SIEGEL R L, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA Cancer J Clin, 2021, 71(3): 209-249. DOI: 10.3322/caac.21660.

[2] KUKURBA K R, MONTGOMERY S B. RNA sequencing and analysis[J]. Cold Spring Harb Protoc, 2015, 2015(11): 951-969. DOI: 10.1101/pdb.top084970.

[3] LUCCHINETTI E, ZAUGG M. RNA sequencing[J]. Anesthesiology, 2020, 133(5): 976-978. DOI: 10.1097/aln.0000000000003524.

[4] LI Jie, CHEN Chaomei. CiteSpace: Technology Text Mining and Visualization[M]. Beijing: Capital University of Economics and Business Press, 2016.

[5] SABE M, PILLINGER T, KAISER S, et al. Half a century of research on antipsychotics and schizophrenia: a scientometric study of hotspots, nodes, bursts, and trends[J]. Neurosci Biobehav Rev, 2022, 136: 104608. DOI: 10.1016/j.neubiorev.2022.104608.

[6] Qiu Junping. Bibliometrics[M]. 2nd ed. Beijing: Science Press, 2019.

[7] LIU L M, XIONG D D, LIN P, et al. DNA topoisomerase 1 and 2A function as oncogenes in liver cancer and may be direct targets of nitidine chloride[J]. Int J Oncol, 2018, 53(5): 1897-1912. DOI: 10.3892/ijo.2018.4531.

[8] SUN Y F, WU L, ZHONG Y, et al. Single-cell landscape of the ecosystem in early-relapse hepatocellular carcinoma[J]. Cell, 2021, 184(2): 404-421.e16. DOI: 10.1016/j.cell.2020.11.041.

[9] ZHANG B J, FAN T. Knowledge structure and emerging trends in the application of deep learning in genetics research: a bibliometric analysis[2000-2021][J]. Front Genet, 2022, 13: 951939. DOI: 10.3389/fgene.2022.951939.

[10] SHIGETA K, MATSUI A, KIKUCHI H, et al. Regorafenib combined with PD1 blockade increases CD8 T-cell infiltration by inducing CXCL10 expression in hepatocellular carcinoma[J]. J Immunother Cancer, 2020, 8(2): e001435. DOI: 10.1136/jitc-2020-001435.

[11] ZHANG J Q, HAN C, UNGERLEIDER N, et al. A transforming growth factor-β and H19 signaling axis in tumor-initiating hepatocytes that regulates hepatic carcinogenesis[J]. Hepatology, 2019, 69(4): 1549-1563. DOI: 10.1002/hep.30153.

[12] NAGY Á, LÁNCZKY A, MENYHÁRT O, et al. Validation of miRNA prognostic power in hepatocellular carcinoma using expression data of independent datasets[J]. Sci Rep, 2018, 8(1): 9227. DOI: 10.1038/s41598-018-27521-y.

[13] THOMPSON S M, SUMAN G, TORBENSON M S, et al. PSMA as a theranostic target in hepatocellular carcinoma: immunohistochemistry and 68 Ga-PSMA-11 PET using cyclotron-produced 68 Ga[J]. Hepatol Commun, 2022, 6(5): 1172-1185. DOI: 10.1002/hep4.1861.

[14] LAU C C, SUN T T, CHING A K, et al. Viral-human chimeric transcript predisposes risk to liver cancer development and progression[J]. Cancer Cell, 2014, 25(3): 335-349. DOI: 10.1016/j.ccr.2014.01.030.

[15] LONG J Y, ZHANG L, WAN X S, et al. A four-gene-based prognostic model predicts overall survival in patients with hepatocellular carcinoma[J]. J Cell Mol Med, 2018, 22(12): 5928-5938. DOI: 10.1111/jcmm.13863.

[16] SHEN H, LIU B Q, XU J J, et al. Circular RNAs: characteristics, biogenesis, mechanisms and functions in liver cancer[J]. J Hematol Oncol, 2021, 14(1): 134. DOI: 10.1186/s13045-021-01145-8.

[17] NAKAGAWA H, FUJITA M, FUJIMOTO A. Genome sequencing analysis of liver cancer for precision medicine[J]. Semin Cancer Biol, 2019, 55: 120-127. DOI: 10.1016/j.semcancer.2018.03.004.

[18] CHEN G, NING B T, SHI T L. Single-cell RNA-seq technologies and related computational data analysis[J]. Front Genet, 2019, 10: 317. DOI: 10.3389/fgene.2019.00317.

[19] KE X, WU H, CHEN Y X, et al. Individualized pathway activity algorithm identifies oncogenic pathways in pan-cancer analysis[J]. EBioMedicine, 2022, 79: 104014. DOI: 10.1016/j.ebiom.2022.104014.

[20] YU A T, BERASAIN C, BHATIA S, et al. PHAROH lncRNA regulates myc translation in hepatocellular carcinoma via sequestering TIAR[J]. Elife, 2021, 10: e68263. DOI: 10.7554/eLife.68263.

[21] CHEN C M, SONG M. Visualizing a field of research: a methodology of systematic scientometric reviews[J]. PLoS One, 2019, 14(10): e0223994. DOI: 10.1371/journal.pone.0223994.

(Received: August 16, 2023; Revised: September 14, 2023)
(Editor: CAO Xinyang)

Note: Figure translations are in progress. See original paper for figures.