INTEGRATION METHODS FOR MULTI-OMICS DATA: A BIBLIOMETRIC ANALYSIS
ÇOKLU-OMİK VERİLER İÇİN BİRLEŞTİRME YÖNTEMLERİ: BİBLİYOMETRİK BİR ANALİZ

Ahu CEPHE[1]

[1]Erciyes University Rectorate, Institutional Data Management and Analytics Unit, Kayseri, Türkiye

ABSTRACT

The primary objective of this study is to conduct a bibliometric analysis of multi-omics publications from 2013 to 2023 in the Web of Science database, utilizing visual mapping techniques. Network maps were generated using Biblioshiny and VOSviewer software to illustrate the published trends over the investigated period. A comprehensive examination of 714 articles from 298 journals aimed to unveil the intellectual structure and emerging trends in the multi-omics field. For this purpose, co-authorship, bibliographic coupling, and co-occurrence analyses were conducted for country, institution, source, author, and keyword productivity. During the study period, China emerged as the leading contributor to multi-omics publications, while the USA secured the highest number of multi-omics citations. The most frequently occurring terms in the author's keywords were identified as "multi-omics," "data-integration," and "metabolomics." The study also determined "Bioinformatics Briefings" as both the most relevant source and the most cited. Temporal analysis indicated a noteworthy increase in publications from 2013 to 2022. This observation suggests a significant rise in interest and research activity in the multi-omics field during the specified period.

Keywords: Bibliometric analysis, co-authorship, data integration, gene expression, multi-omics.

ÖZ

Bu çalışmanın temel amacı, görsel haritalama tekniklerini kullanarak Web of Science veritabanında 2013-2023 yılları arasında çoklu-omik yayınların bibliyometrik analizini yapmaktır. İncelenen dönemdeki yayınların eğilimleri göstermek için Biblioshiny ve VOSviewer yazılımları kullanılarak ağ haritaları oluşturuldu. 298 dergiden 714 makalenin kapsamlı bir incelemesi, çoklu-omik alanındaki entelektüel yapıyı ve oluşan eğilimleri ortaya çıkarmayı amaçladı. Bu amaçla, ülke, kurum, kaynak, yazar ve anahtar kelime üretkenliği için ortak-yazarlık, bibliyografik bağlantı ve eş-dizimlilik analizleri yapıldı. Çalışma süresince, Çin çoklu-omik yayınlara önde gelen katkıda bulunan ülke olarak ortaya çıkarken, ABD en fazla çoklu-omik yayını atıf sayısına ulaştı. Yazarın anahtar kelimelerinde en sık görülen terimler "çoklu-omik", "veri entegrasyonu" ve "metabolomik" olarak belirlendi. Çalışma ayrıca "Bioinformatics Briefings" dergisini hem en ilgili kaynak hem de en çok atıf yapılan kaynak olarak belirledi. Zamansal analiz, 2013'dan 2022'ye kadar yayınlarda kayda değer bir artış olduğunu gösterdi. Bu gözlem, belirtilen dönemde çoklu-omik alanında ilgi ve araştırma faaliyetlerinde önemli bir artış olduğunu gösteriyor.

Anahtar kelimeler: Bibliyometrik analiz, ortak yazarlık, veri entegrasyonu, gen ifadesi, çoklu-omikler.

Corresponding Author: Dr. Ahu CEPHE, ahucephe@erciyes.edu.tr, 0000-0001-9374-4495, Erciyes University Rectorate, Institutional Data Management and Analytics Unit, Kayseri, Türkiye.

INTRODUCTION

In recent years, researchers have evaluated the diagnosis and treatment of diseases based on the results of single omics data analysis, such as metabolomics, genomics, proteomics, and transcriptomics. However, there are uncertainties regarding their accuracy as the results based on these single omics data can only identify some of the diseases.[1] At this point, single-omics data is insufficient, and multi-omics data is needed for diagnosis and personalized treatments of diseases.[2] Many studies in the world literature use integration methods to analyze multi-omics data, and the number of these studies is increasing daily.

These integration studies can produce more precise solutions to the problems of complex diseases. Such integrated approaches provide promising advances towards a systematic and holistic understanding of biology. Because multi-omics data helps bridge the gap between genotype and phenotype, it is possible to understand and evaluate molecular interactions.[3,4]

Various methodologies for integrating multi-omics methods have emerged depending on the type, quality, and data availability. These methodologies are grouped under Bayesian, network, correlation, similarity, multivariate, and fusion.[4] Cavill et al. classified integration methods under the headings conceptual, statistical, and model-based integration.[5] Lin and Lane focuses on model-based integration, concatenate-based integration, and transformation-based integration.[6] Ritchie et al. evaluated integration methods under the headings of multi-staged and meta-dimensional analyses.[3] Ivanisevic and Sewduthdiscussed integration studies under conceptual integration, model-based integration, statistical integration, networks, and pathway data integration.[7]

This article used bibliometric analysis in studies on multi-omics integration methods. Bibliometric analysis is a popular method for many purposes, such as determining trends in article and journal performances and revealing research components. The availability and accessibility of bibliometric software, such as VOSviewer[8] (www.vosviewer.com) and CiteSpace[9] (https://citespace.podia.com/), and scientific databases, such as Scopus (https://www.scopus.com/sources.uri) and Web of Science (WoS) (https://www.webofscience.com/wos/woscc/basic-search), make it easy to carry out these analyses. The convenience provided by these analyses offers advantages, such as identifying prominent research topics in academic studies, observing trends in research models, following the change of studies in the literature over time, choosing journals, and evaluating research outputs quantitatively.[10,11] It is generally used to assess existing knowledge in research areas and to identify new research topics. Our study aimed to investigate and assess the current status of integration methods used in multi-omics data between 2013 and 2023 in the literature using bibliometric analysis.

MATERIAL AND METHODS

Database Creation

In this study, only the WoS was used as the data source. WoS was chosen for its inclusion of high-quality, peer-reviewed publications, standardized and reliable metadata, and advanced citation analysis tools. While other databases such as Scopus, PubMed, and Dimensions also offer valuable bibliographic content, differences in indexing criteria, journal coverage, and citation metrics may lead to data heterogeneity and inconsistencies when combined. To ensure methodological consistency, data integrity, and comparability with similar studies, the analysis was limited to WoS. This decision was also influenced by practical factors, including technical constraints, time efficiency, and institutional access. The database searches were conducted on January 16, 2023 using the WoS database. The search query utilized in the examination of scholarly publications is explicated as follows: ((TI= ("multi-omics" AND ("integration" OR "integrated" OR "combination" OR "integrating" OR "combining" OR "fusion")) OR AK= (("multi-omics") AND ("data integration" OR "integration"))) OR (QMTS= ("MULTI-OMICS INTEGRATION ANALYSIS"))). The obtained results from the literature were exported to a Plain Text File (.txt). At this stage, each record's full record and cited references were selected.

Bibliometric Analysis

In conducting comprehensive analyses encompassing research trends, content evaluation, and keyword scrutiny, we employed biblioshiny version 4.1.4 as a web interface facilitating the utilization of the bibliometrix R package.[12] Additionally, VOSviewer software version 1.6.20 was utilized for further examination.[8] The search query in the WoS database yielded 964 literature records about the integration of multi-omics data spanning 2013 to 2023. Among these, 185 records falling under diverse document types, such as meeting abstracts, early access, editorial materials, proceeding papers, book chapters, letters, data papers, and notes, were subjected to filtration. Subsequently, out of the identified 779 articles and review articles, two were excluded due to being in languages other than English. Although the keywords in the search query created in the 'Database Creation' section were selected based on their usage in various parts of the articles, 63 articles and review articles that did not include new algorithms or applications of multi-omics integration methods in terms of content were excluded from the scope of the study, and the analyses were carried out on 714 of the 777 publications. Following this meticulous filtration process, 714 articles and review articles in English were included in the subsequent bibliometric analyses, as illustrated in Figure 1.a. An average citation frequency is 18.62 times per article, and annual growth rate is 57.09%. Figure 1.b graphically depicts the temporal evolution of article numbers and citations, indicating a notable upsurge, particularly in the peak values observed in 2022-2023. An insightful observation is that 636 articles were published between 2019 and 2023, constituting 89.08% of the overall article count.

To explore international collaboration patterns, co-authorship country analysis was performed. This method identifies leading countries, active participants, regional/global clusters, and nations involved in strategic collaborations despite low output. In the network, nodes represent countries, edges show collaboration counts, link strength indicates shared

publications, and centrality reflects a country's role in the global scientific network. Bibliographic coupling with source reveals citation-based links between journals that frequently cite the same sources. This analysis helps identify journals with similar literature bases, thematic clusters, and key outlets in the field, guiding researchers toward suitable publication venues and highlighting the concentration of relevant studies. Keyword co-occurrence network analysis identifies thematic structures and topic relationships by linking keywords that frequently appear together. It reveals major research themes, emerging trends, interdisciplinary links, and potential gaps. In the net-work, clusters show related topics, node size reflects keyword frequency, and link strength indicates co-occurrence in publications.

RESULTS
Quantitative Analysis of Publication

After applying year and language filters, a total of 714 multi-omics studies, comprising 596 articles and 118 review articles, were incorporated into this bibliometric analysis (Figure 1.c). The annual publication figures from 2013 to 2023 are depicted in Figure 1.b. The annual publication count was relatively modest during the initial phase, spanning 2013 to 2015. Subsequently, from 2015 to 2017, the publication rate remained nearly constant, averaging 13.33 publications annually. A consistent upswing in multi-omics studies was observed from 2017 to 2022, with the culmination of publications occurring in 2022. Notably, there was a decline in publications in 2023, attributed to the incomplete representation of all studies conducted in that year within the WoS database.

a
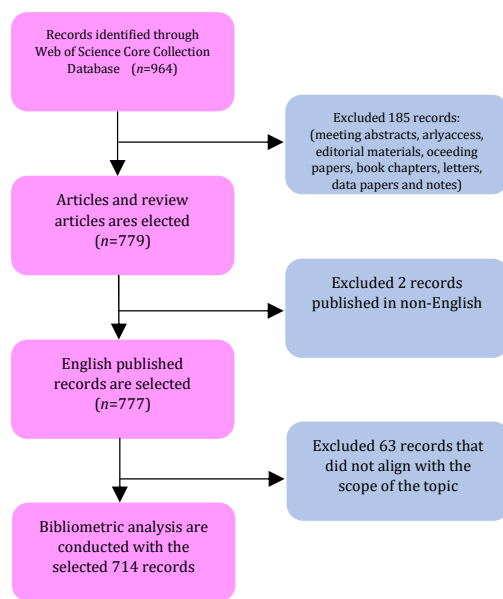
c

| Description	Results
Timespan	2013:2023
Sources (Journals, Books, etc)	298
Documents	714
Annual Growth Rate %	57.09
Document Average Age	2.94
Average citations per doc	18.62
References	40623
DOCUMENT CONTENTS	
Keywords Plus (ID)	2170
Author's Keywords (DE)	1744
AUTHORS	
Authors	4884
Authors of single-authored docs	12
AUTHORS COLLABORATION	
Single-authored docs	12
Co-Authors per Doc	8.05
International co-authorships %	30.39
DOCUMENT TYPES	
article	559
article; book chapter	13
article; data paper	1
article; early access	17
article; proceedings paper	6
review	115
review; early access	3

b
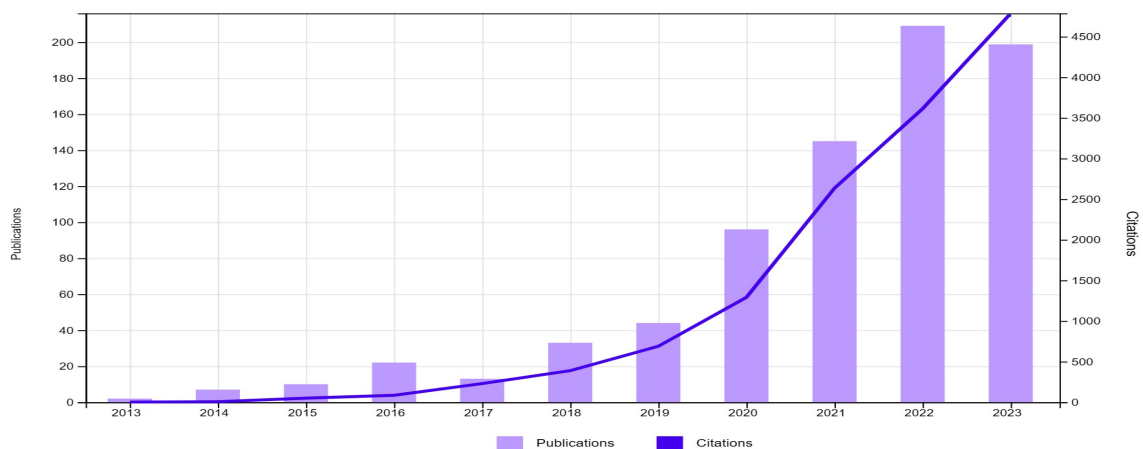
Figure 1. (a) Bibliometric analysis flow diagram of the study selection process, (b)Temporal evolution of article numbers and citations in the field of multi-omics, (c) Main data information.

Countries and Institutions Analysis

A total of 66 countries and 1153 institutions were encompassed in multi-omics research. The top 10 countries exhibit a global distribution, spanning Asia, America, Europe, and Oceania. China leads with the highest number of publications (n=297), followed by the USA (n=194) and Germany (n=57).

A co-authorship country analysis was conducted to elucidate international collaborations. The total strength of each country and its co-authorship links with others were calculated (Figure 2.a). In the network visualization, each circle represents an author's country, with font size denoting the frequency of collaborations (Figure 2.b). Lines connecting countries depict collaboration strength, while the node indicates the number of collaborations from each country. Countries of the same color share a similar research area.

The analysis revealed 5 clusters, with a total of 198 links and 533 total link strength. In cluster 1 (red), Germany (n=57), Belgium (n=11), England (n=49), France (n=33), Ireland (n=5), Italy (n=30), Netherlands (n=29), and Switzerland (n=13) exhibit significant co-authorship. In cluster 2 (green), Denmark (n=16), Finland (n=9), Iran (n=7), Norway (n=10), Poland (n=5), Portugal (n=6), and Sweden (n=18) demonstrate close cooperation. In cluster 3 (blue), Canada (n=43), Japan (n=12), China (n=297), Singapore (n=7), South Korea (n=27), Taiwan (n=5), and USA (n=194) are deeply linked in multi-omics research. In cluster 4 (yellow), Egypt (n=5), Greece (n=7), Saudi Arabia (n=5), and Spain (n=28) collaborate closely. In cluster 5 (purple), Australia (n=25), Austria (n=8), Brazil (n=13), and India (n=34) exhibit significant co-authorship.

Notably, the USA ranks first with 165 total link strength (Figure 2.a). China is the leading country in producing publications on the subject of multi-omics. However, when examining the citations of these publications, the USA ranks first in citations, followed by China and Germany.

Regarding the most pertinent affiliations determined by the corresponding authors of articles, the Helmholtz Association (Germany) and the University of California System (USA) leads with the highest number of articles (n=50), as illustrated in Figure 2.c. Following closely is
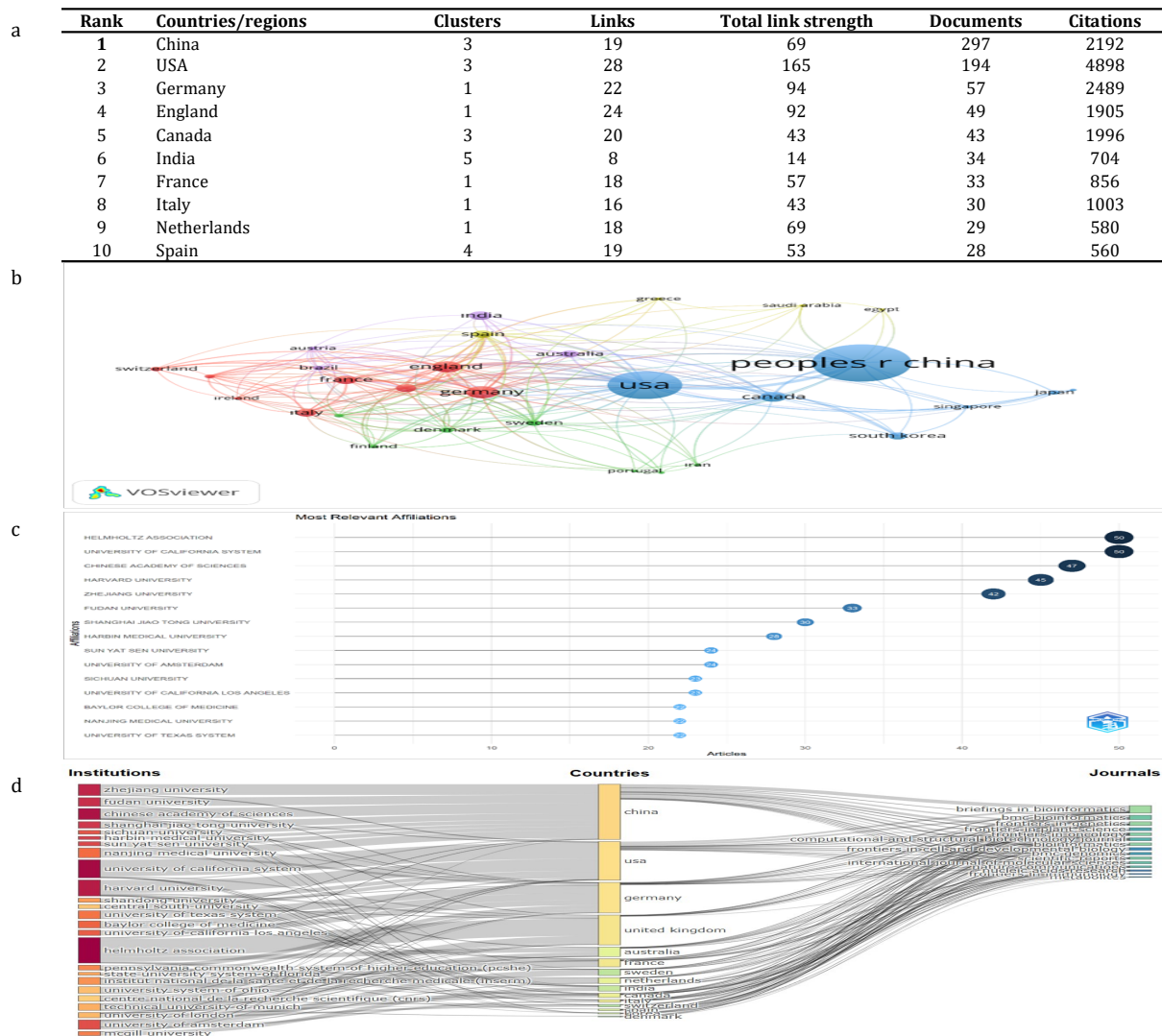
a

Rank	Countries/regions	Clusters	Links	Total link strength	Documents	Citations
1	China	3	19	69	297	2192
2	USA	3	28	165	194	4898
3	Germany	1	22	94	57	2489
4	England	1	24	92	49	1905
5	Canada	3	20	43	43	1996
6	India	5	8	14	34	704
7	France	1	18	57	33	856
8	Italy	1	16	43	30	1003
9	Netherlands	1	18	69	29	580
10	Spain	4	19	53	28	560

b

c

d

Figure 2. (a)The topmost 10 strong co-authorship-linked document-productive countries/regions, (b) Map of visualization of countries on research of multi-omics, (c) Most relevant institutions, (d) Three-field plot showing the network between institutions (left), countries (middle), and journals (right).

the Chinese Academy of Sciences, which contributes 47 articles, and Harvard University, with 45 articles.

Figure 2.d depicts a three-field plot utilizing the Sankey chart, elucidating the interplay among countries, institutions, and journals. The height of the rectangular nodes corresponds to the frequency of a specific country, institution, or journal within the collaborative network. Meanwhile, the width of the lines connecting nodes is proportionate to the number of connections. This analysis evaluated 25 affiliations, 15 countries, and 20 journals. Germany had the highest diversity in affiliation information, with 14 different affiliations, followed by China with 12, the USA with 11, and the United Kingdom with 10. The journal receiving the most publications from various countries was "Frontiers in Oncology", featuring articles from 16 different countries. This was followed by "Briefings in Bioinformatics" and "Frontiers in Genetics", each with publications from 15 different countries.

Bibliographic Coupling with Sources

A total of 298 sources were generated from research articles, and the full counting method was applied with a minimum threshold of 5. Only 33 sources met these criteria. For each of the 33 sources, the total strength of bibliographic coupling links with other sources was calculated (refer to Figure 3.a, Figure 3.b). The analysis revealed 526 links and a total link strength of 23942, forming 2 clusters with 33 items. The first cluster

comprised 21 elements, while the second had 12 elements.

Sources in the first cluster include Biomolecules (n=8), BMC Genomics (n=11), Cells (n=5), Clinical Epigenetics (n=5), Communications Biology (n=6), Frontiers in Cell and Developmental Biology (n=10), Frontiers in Immunology (n=10), Frontiers in Pharmacology (n=6), Frontiers in Plant Science (n=12), Genome Biology (n=6), International Journal of Molecular Sciences (n=16), iScience (n=5), Journal of Translational Medicine (n=8), Metabolites (n=11), Microbiome (n=5), Molecular Omics (n=6), Nature Communications (n=13), Nucleic Acids Research (n=10), OMICS: A Journal of Integrative Biology (n=8), Plos One (n=7), and Scientific Reports (n=18).

In the second cluster, resources include Bioinformatics (n=19), BMC Bioinformatics (n=16), Briefings in Bioinformatics (n=31), Cancers (n=7), Computational and Structural Biotechnology Journal (n=12), Computers in Biology and Medicine (n=10), Current Bioinformatics (n=5), Frontiers in Genetics (n=29), Frontiers in Oncology (n=18), Genes (n=6), Journal of Biomedical Informatics (n=5), and Plos Computational Biology (n=6).

The source with the highest total link strength is "Briefings in Bioinformatics," followed by "Frontiers in Genetics." This indicates strong collaboration between these two source journals in the publication of scientific
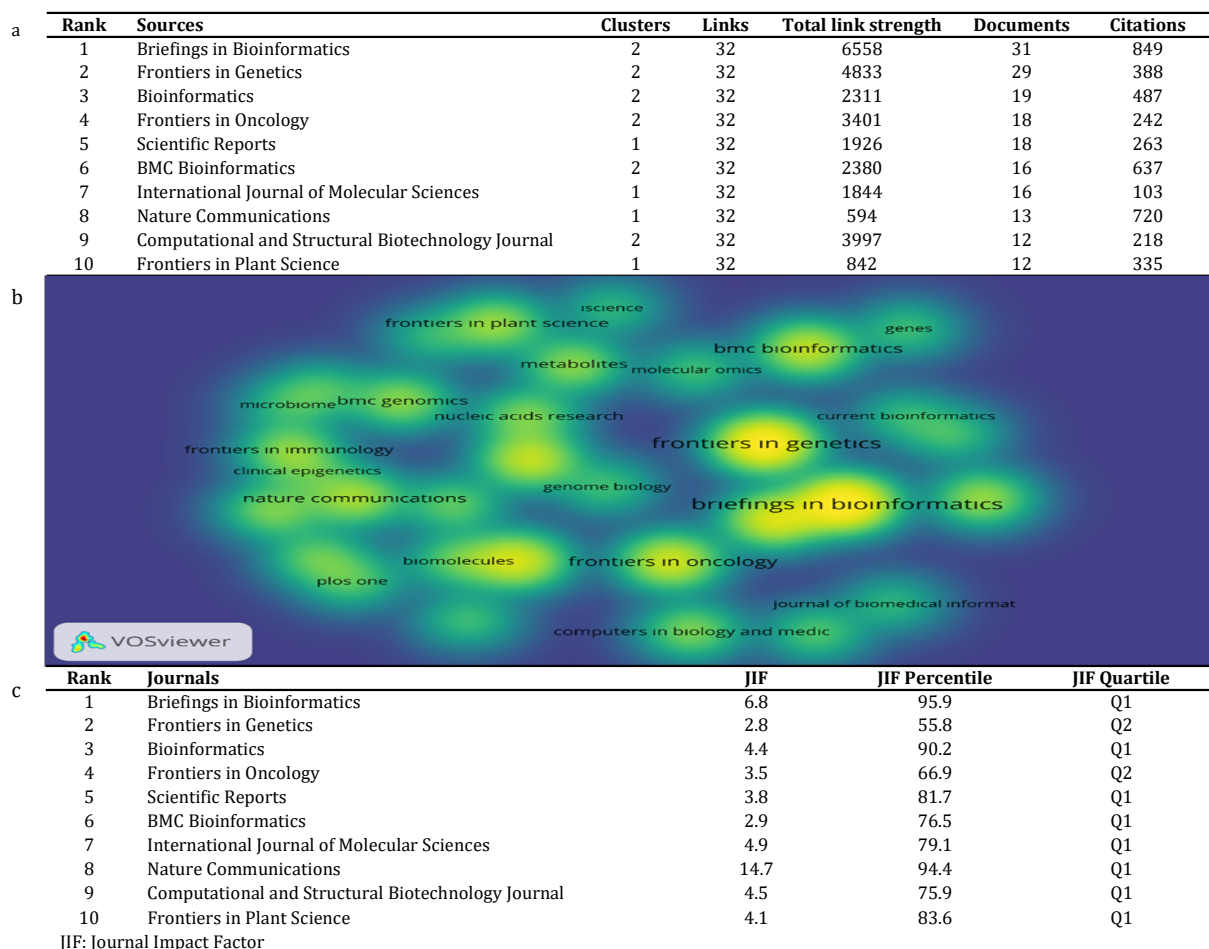
a

Rank	Sources	Clusters	Links	Total link strength	Documents	Citations
1	Briefings in Bioinformatics	2	32	6558	31	849
2	Frontiers in Genetics	2	32	4833	29	388
3	Bioinformatics	2	32	2311	19	487
4	Frontiers in Oncology	2	32	3401	18	242
5	Scientific Reports	1	32	1926	18	263
6	BMC Bioinformatics	2	32	2380	16	637
7	International Journal of Molecular Sciences	1	32	1844	16	103
8	Nature Communications	1	32	594	13	720
9	Computational and Structural Biotechnology Journal	2	32	3997	12	218
10	Frontiers in Plant Science	1	32	842	12	335

b

c

Rank	Journals	JIF	JIF Percentile	JIF Quartile
1	Briefings in Bioinformatics	6.8	95.9	Q1
2	Frontiers in Genetics	2.8	55.8	Q2
3	Bioinformatics	4.4	90.2	Q1
4	Frontiers in Oncology	3.5	66.9	Q2
5	Scientific Reports	3.8	81.7	Q1
6	BMC Bioinformatics	2.9	76.5	Q1
7	International Journal of Molecular Sciences	4.9	79.1	Q1
8	Nature Communications	14.7	94.4	Q1
9	Computational and Structural Biotechnology Journal	4.5	75.9	Q1
10	Frontiers in Plant Science	4.1	83.6	Q1

JIF: Journal Impact Factor

Figure 3. (a) The topmost 10 strong bibliographic coupling with source, (b) Bibliographic coupling with sources, (c) Top 10 journals with the most papers.

articles.

Within the Top 10 journals, "Briefings in Bioinformatics" stands out as the foremost authoritative review journal in biochemical research methods, boasting the highest Journal Impact Factor (JIF) in 2023 as indicated in Figure 3.c. "Nature Communications" had the highest impact value at 14.7. In 2023, its JIF was 94.4, making it the journal with the second-highest JIF after "Briefings in Bioinformatics".

Figure 4.a presents the top 20 most published journals in the field of multi-omics, while Figure 4.b illustrates the annual total number of articles published by the top 5 most productive journals. Topping the list of most published journals is the "Briefings in Bioinformatics" journal, boasting 31 articles, closely followed by "Frontiers in Genetics" with 29 articles. An examination of the total number of citations for relevant publications reveals that "Briefings in Bioinformatics" ranks first with 849 citations, followed by "Nature Communications" with 720 citations and "BMC Bioinformatics" with 637 citations (Figure 3.a).

Figure 4.c illustrates the inaugural academic articles on multi-omics within the Bradford's Law graph district. Journals in this central region are considered top-tier publications in the multi-omics field, serving as the primary outlets for most articles on multi-omics research. This chart holds significance for researchers seeking to pinpoint the most productive journals in multi-omics.

Highly Cited Articles and Most Influential Authors

Table 1 presents the first five highly cited research articles. Leading the citations in the field of multi-omics is the article titled "Multi-omics Data Integration, Interpretation, and Its Application" by Subramanian et al.[4] Published in 2020, this article has amassed an impressive 443 citations within a span of 3 years.

In Figure 5, the most prolific authors in the field of multi-omics are depicted based on the number of publications attributed to each author. The magnitude of the circles in the visualization signifies the number of publications authored by an individual in the respective year, while the intensity of the circles represents the count of citations garnered by the author during that same year. As the number of publications and citations rise, the size and darkness of the circles proportionally increase. Notably, authors such as Li, X., Zhang, Y., Li, J., and Kim, D. initiated their studies on the subject in 2014, while other prolific authors exhibited increased activity, particularly after 2019. This underscores that multi-omics studies have gained prominence over the
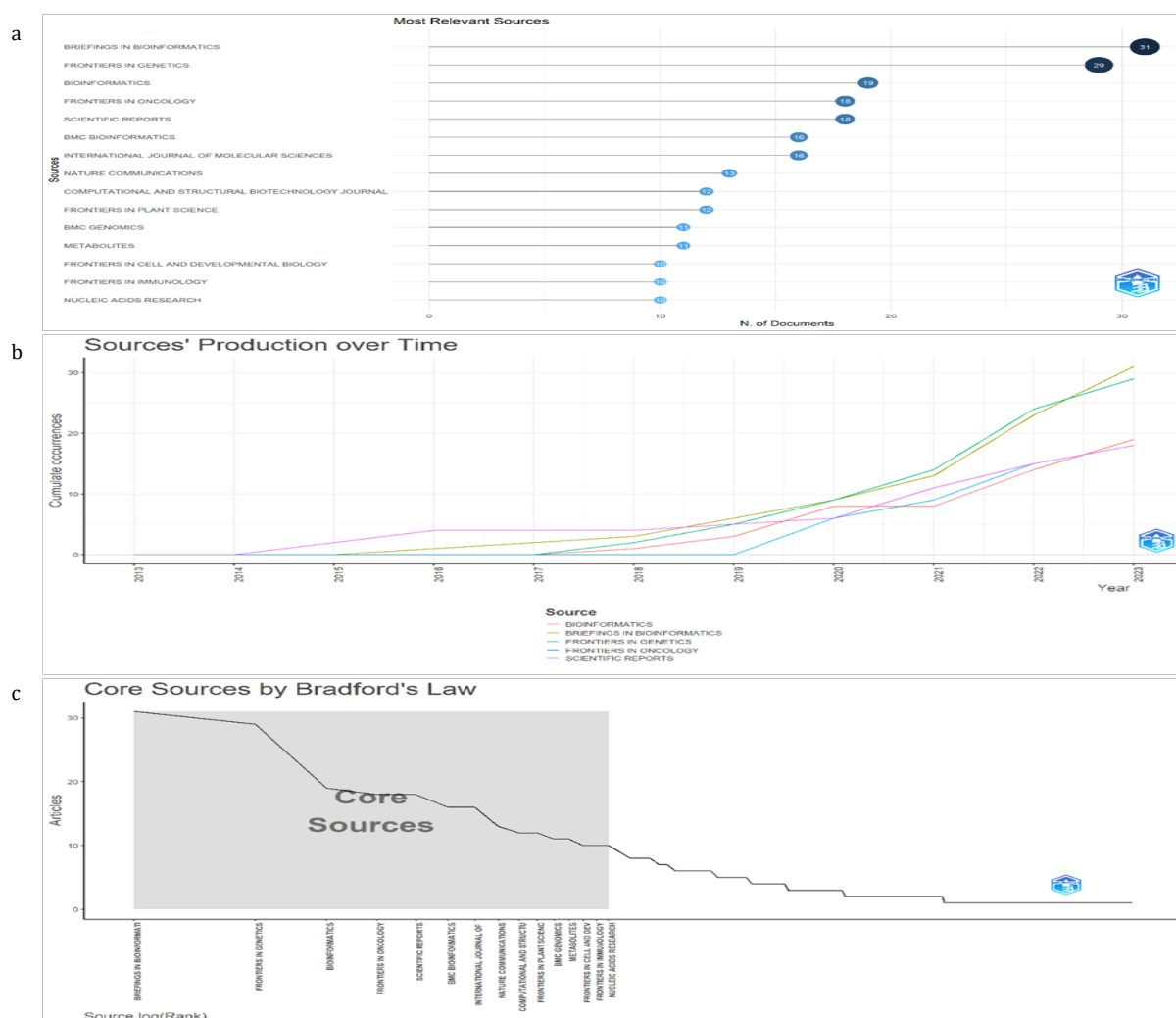
Figure 4. Most relevant sources (a) Most relevant sources, (b) Sources' Production over Time, (c) Core Sources by Bradford's Law.

Table 1. Highly cited articles.[4,13,14,15,16]

Rank	Title	Citation	Total Citations	Total Citations per Year	Normalized Total Citations
1	Multi-omics Data Integration, Interpretation, and Its Application	Subramanian et al., 2020[4]	443	88.60	12.72
2	Multi-Omics Factor Analysis-a framework for unsupervised integration of multi-omics data sets	Argelaguet et al., 2018[13]	424	60.57	5.85
3	Deep Learning-Based Multi-Omics Integration Robustly Predicts Survival in Liver Cancer	Chaudharyl et al., 2018[14]	283	40.43	3.90
4	Using MetaboAnalyst 5.0 for LC-HRMS spectra processing, multi-omics integration and covariate adjustment of global metabolomics data	Pang et al., 2022[15]	359	119.67	39.33
5	Systems Biology and Multi-Omics Integration: Viewpoints from the Metabolomics Research Community	Pinu et al., 2019[16]	316	52.67	8.65

last five years, establishing themselves as a current and dynamic research field. Among the most productive authors, Wang, J. (Zhejiang University, China) leads with 16 articles, followed by Li, Y. (Beijing Institute of Microbiology and Epidemiology, China) with 14 articles, and Li, X. (Henan University of Science and Technology, China) with 12 articles.

expression", "glioma", "hepatocellular carcinoma", "immunotherapy", "integrative analysis", "lung cancer", "machine learning", "methylation", "multi-omics analysis", "multi-omics integration", "ovarian cancer", "pan-cancer", "prognosis", "prognostic model", "single-cell", "survival prediction", and "TCGA". Cluster 2 (green) "autoencoder", "cancer", "cancer subtyping",
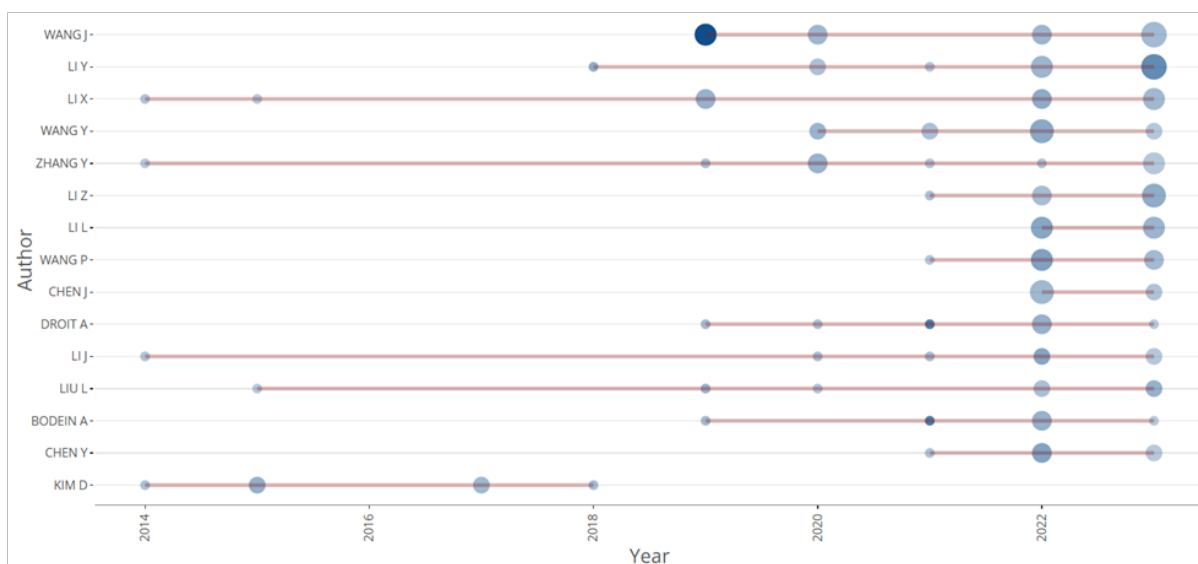
Figure 5. Production of the most productive authors over time.

Keyword Analysis

Keyword analysis based on author's keywords is crucial for gaining insights into core topics, focus areas, and trends within a research field. Such analyses enable researchers to comprehend the most discussed topics and key concepts swiftly. The word cloud presented in Figure 6.a highlights the most frequently used keywords in the relevant field, prominently featuring terms such as "multi-omics," "data integration," and "metabolomics". This suggests a prevalent utilization of metabolomics data in the context of multi-omics studies. Figure 6.b shows the result of a network of keywords that co-occur in multi-omics. The 73 most commonly used words in multi-omics are divided into 7 clusters. Cluster 1 (red) "aging", "bioinformatics", "biomarker", "breast cancer", "colon cancer", "colorectal cancer", "copy number variation", "DNA methylation", "gene

"data integration", "deep learning", "dimension reduction", "feature selection", "gwas", "multi-omics data", "multi-omics data integration", "multiple kernel learning", and "survival analysis". Cluster 3 (blue) "abiotic stress", "genomics", "gut microbiome", "metabolomics", "metagenomics", "metatranscriptomics", "microbiome", "omics integration", "phenomics", "proteomics", and "transcriptomics". Cluster 4 (yellow) "artificial intelligence", "big data", "epigenetics", "epigenomics", "integration", "multi-omics", "network", "omics", "precision medicine", and "rna- seq". Cluster 5 (purple) is "alzheimer's disease", "biomarkers", "covid-19", "integration", "lipidomics", "personalized medicine", and "systems biology". Cluster 6 (turquois) is "data analysis", "multi-omics integration", "network analysis", "pathway analysis", and "transcriptomic". Cluster 7

(orange) is "metabolome", "proteome", and "transcriptome".

The keyword co-occurrence network analysis reveals that multi-omics research is characterized by its multidisciplinary nature. Studies within the blue cluster are presumed to be predominantly conducted by researchers in the field of biology. In contrast, researchers likely lead those in the red cluster in the health sciences. Notably, the involvement of statisticians and data scientists is evident in studies associated with keywords highlighted in yellow and green ranks.

According to trend topic analysis, in 2015, terms such as "dimension reduction" and "data analysis" dominated the research, while in 2018, the term "network analysis" gained prominence. As of 2019, "multi-omics" and "data integration" started manifesting themselves. After 2020, sub-keywords under the main cluster terms "multi-omics" and "data integration" began to emerge. Therefore, it has been determined that 2019 marks a turning point in the bibliometric analysis of the multi-omics field, with a shift towards a focus on multi-omics data integration (Figure 6.c).

DISCUSSION

This study conducted bibliometric analyses by compiling research articles and review articles published in the WoS journal scanning index related to multi-omic studies between 2013 and 2023. Using Biblioshiny and VOSviewer, this study aimed to comprehend the current literature on multi-omic studies and present the global and accessible overview of studies conducted until January 1, 2024. Due to the increased advancements in omics technologies in recent years, it was observed that studies that showed a low trend between 2013 and 2015 reached an average of 13.33 articles between 2015 and 2017. However, the highest increase was recognized in the year 2022. The integration of data obtained in the field of multi-omics has become necessary, with an increasing number of studies conducted each year to understand the complex interactions between the data. Therefore, it is anticipated that there will be a substantial number of articles and review articles in this field.

The analysis of international collaboration clusters reveals not only geographic patterns but also distinct thematic focuses across regions in multi-omics research.

a

b

c
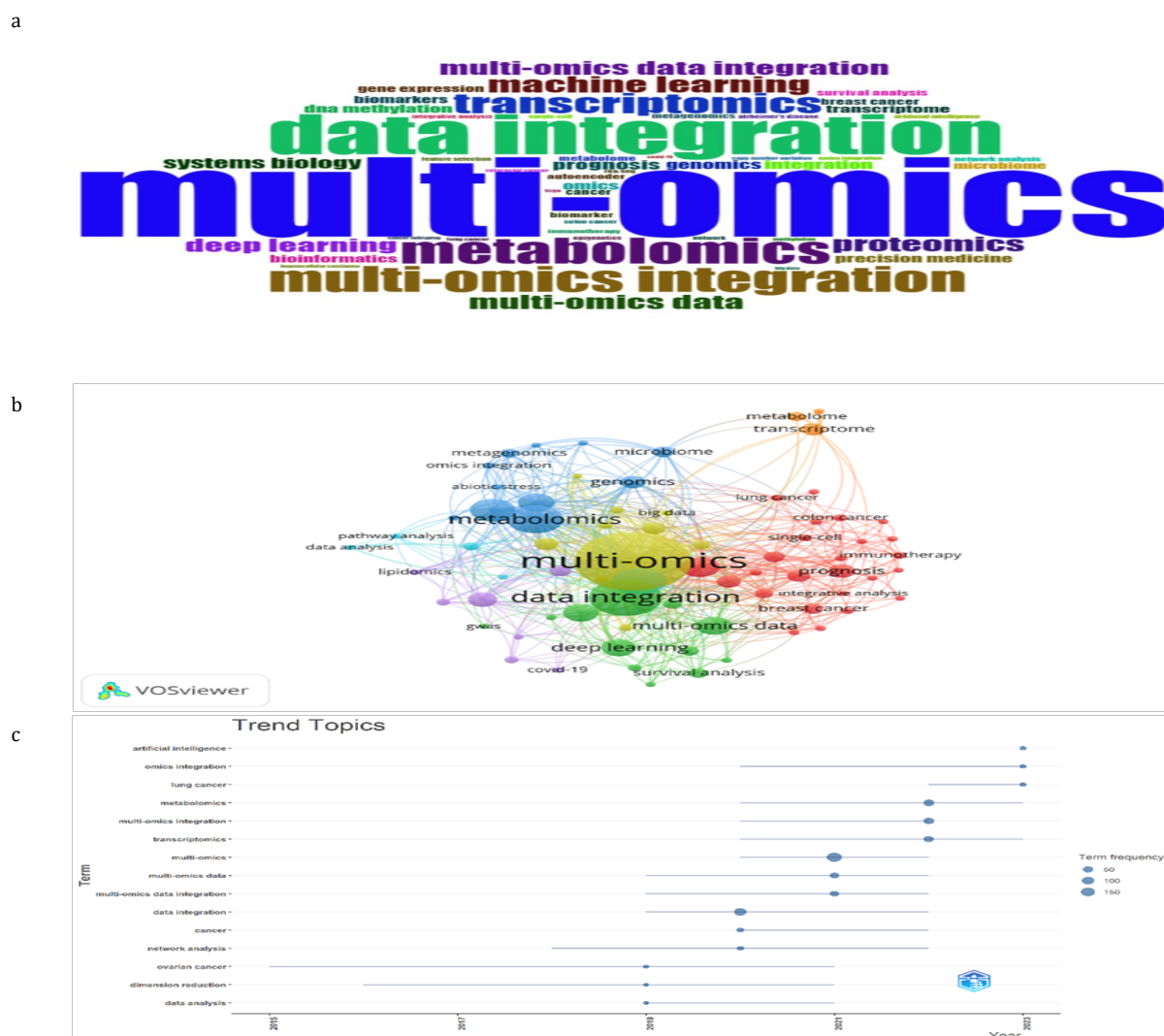
Figure 6. (a) Word cloud of the most frequent terms in the author's keywords of selected articles, (b) Keyword co-occurrence network in multi-omics research, (c) Trend topics based on author's keywords over time.

Cluster 1 (red), consisting of European countries such as Germany, England, France, the Netherlands, and Italy, is characterized by leadership in interdisciplinary projects primarily focused on cancer bioinformatics, systems biology, and translational medicine. Cluster 3 (blue), which includes the USA, China, Canada, Japan, and South Korea, plays a leading role in methodological innovations, particularly in artificial intelligence and machine learning approaches for multi-omics data integration. Cluster 2 (green) encompasses Nordic countries and some emerging nations, contributing significantly to population-based epidemiological and neurodegenerative disease studies. Cluster 4 (yellow), composed of Spain, Egypt, Saudi Arabia, and Greece, focuses on region-specific health issues, particularly metabolomics and microbiome research. Finally, cluster 5 (purple), including India, Australia, Brazil, and Austria, demonstrates active participation in diverse areas such as plant genomics, metabolic diseases, and bioinformatics infrastructure development. These thematic and regional differences within the collaboration networks reflect critical dynamics shaping the scientific advancement and translational impact of multi-omics research globally.

Within this study's scope, 714 studies were retrieved from the WoS database between 2013 and 2023. Since 2013, the number of articles published in this field has increased geometrically, with an annual growth rate of 57.09%. This increase is expected, given the recent advances in omics technologies over the last 15 years. The top 10 journals with the highest number of publications in the field of multi-omics, along with their WoS categories, were reported as follows: Biochemical Research Methods, Mathematical & Computational Biology, Genetics & Heredity, Biochemical Research Methods, Biotechnology & Applied Microbiology, Oncology, Multidisciplinary Sciences, Biochemistry & Molecular Biology, Chemistry, Multidisciplinary, Biochemistry & Molecular Biology, and Plant Science. The reported research areas encompass multidisciplinary fields such as genetics, mathematics, computational biology, biochemistry, molecular biology, chemistry, and cancer, where multi-omics studies intersect. According to Bradford's Law, the top three journals with the most publications are "Briefings in Bioinformatics", "Frontiers in Genetics", and "Bioinformatics", while the top three journals with the most citations are "Briefings in Bioinformatics", "Nature Communications", and "BMC Bioinformatics". The most highly cited articles were "Multi-omics Data Integration, Interpretation, and Its Application" by Subramanian et al.[4], with 443 citations, and "Multi-Omics Factor Analysis: A Framework for Unsupervised Integration of Multi-Omics Data Sets" by Argelaguet et al.[13], with 424 citations.

Analyzing the graph of the most prolific authors over time, it is evident that from 2014 to 2023, Li, X, Zhang Y, and Li, J consistently published articles in this field. According to the findings, it was determined that there is a turning point in the increase of productivity among authors, notably in 2019.

According to keyword analyses, the terms "multi-omics", "data-integration", and "metabolomics" best represent multi-omics studies. Moreover, it was found that the top three countries and institutions with the highest number of articles are Helmholtz Association (Germany), the University of California System (USA), and the Chinese Academy of Sciences (China). Considering the progression of relevant keywords and technological developments over time, it is presumed that, from an economic, accessibility, and technical standpoint, metabolomic data is the most preferred and prioritized area for researchers. Countries and institutions with most published articles are expected to continue their research in this fundamental area. The most associated connections have been identified in the reported universities.

According to the keyword network analysis results in multi-omics, seven distinct clusters have been identified. These clusters have been categorized as cancer research, data integration studies, network research, omics research, metabolome and transcriptome studies, systems biology, personalized medicine research, and multi-omics research. All these classifications are centered on multi-omics related keywords. The bibliometric analysis conducted in this study serves as a guiding resource for researchers in this field, providing insights not only into potential journals for publishing their work but also in anticipating areas that may require further investigation in the future.

The use of keywords in publications has evolved over the years. In the early years, terms like "multi" or "integration" were rarely used, whereas recent publications frequently feature keywords derived from these terms, such as "multi-omics," "data integration," "omics integration," and "multi-omics integration.

The findings of this bibliometric analysis not only reveal the quantitative growth of multi-omic research but also underscore its increasing relevance in both scientific and clinical domains. As multi-omics approaches integrate genomics, transcriptomics, proteomics, metabolomics, and epigenomics data, they enable a systems-level understanding of complex biological processes. For instance, integrative analyses combining transcriptomic and proteomic data have been successfully used to unravel tumor heterogeneity in breast cancer, allowing researchers to classify subtypes with distinct therapeutic responses.[17] Similarly, metabolomic and microbiome data integration has been employed to identify metabolic signatures associated with type 2 diabetes progression.[18] This integrative perspective holds significant potential for advancing precision medicine, particularly in oncology, metabolic diseases, and neurodegenerative disorders, where molecular heterogeneity plays a key role in disease progression and treatment response.

The observed prominence of cancer research and personalized medicine within keyword clusters indicates that multi-omic strategies are increasingly being applied to identify novel biomarkers, therapeutic targets, and predictive models. A notable example is The Cancer Genome Atlas (TCGA) project, which integrates multi-layered omics data to classify tumors beyond traditional histological methods. Likewise, studies such as Subramanian et al.[4] and Argelaguet et al.[13] provide foundational frameworks for data integration, with methods like Multi-Omics Factor Analysis (MOFA)

demonstrating how unsupervised learning can reveal shared and unique patterns across data types.

Furthermore, the growing emphasis on data integration suggests that future studies will likely explore more advanced computational frameworks and AI-assisted analytic methods to handle high-dimensional multi-omic datasets. Tools such as DeepMO, iClusterPlus, and newer federated learning-based models are beginning to address the challenges of scalability, interpretability, and data privacy in integrative omics research. The global collaboration patterns and leading contributions from institutions such as the Helmholtz Association, the University of California System and Chinese Academy of Sciences also suggest that multi-omic research will continue to expand across borders and disciplines. For example, international consortia like the International Human Epigenome Consortium (IHEC) and the Human Cell Atlas project exemplify the collaborative momentum in this area. Thus, the field is poised to play a central role in shaping the future of biomedical research, with direct implications for translational applications and patient care—ranging from early disease detection and prognosis to therapy optimization and individualized treatment plans.

In light of the findings presented in this bibliometric analysis, several directions can be suggested for future research in the field of multi-omics integration. Despite the increasing volume of publications, there is still a need for standardized frameworks that can handle the heterogeneity and high dimensionality of omics data across different platforms. In particular, the integration of single-cell multi-omics data remains a challenging yet promising area, offering a more granular understanding of cellular mechanisms in complex diseases. Moreover, the development of interpretable artificial intelligence (AI) models that can transparently integrate genomics, transcriptomics, proteomics, metabolomics, and epigenomics is a critical future goal, especially for clinical translation. There is also a growing interest in the application of multi-omics in real-time diagnostics and personalized treatment strategies, particularly in oncology, neurodegenerative disorders, and metabolic syndromes. In addition, ethical and regulatory frameworks around data sharing and privacy in multi-omics research are underexplored and warrant deeper investigation. Therefore, future studies should not only aim to improve technical integration methods but also address interdisciplinary gaps that bridge biology, medicine, computation, and data governance.

The integration of multi-omics data holds great potential for unraveling the multilayered complexity of biological systems. However, the current literature still presents several limitations, highlighting important areas for future research. Key challenges include inconsistencies across omic layers, imbalanced sample sizes, missing data, and the lack of standardized frameworks for biological interpretation. In this context, it is intended that readers—particularly researchers actively working in the field—can draw actionable insights and identify future directions based on the trends revealed in this analysis. For instance, the growing importance of single-cell multi-omics integration, time-series omic analyses, AI-assisted data fusion algorithms, and clinically applicable decision-support tools represent critical topics that warrant deeper exploration. Thus, the study not only summarizes the current landscape but also provides strategic insights to guide and inspire forthcoming scientific efforts.

CONCLUSIONS

To our knowledge, this is the first bibliometric analysis in multi-omics and data integration. In this study, we extensively analyzed the sources, authors, institutions, countries, keywords, and clusters of multi-omics and data integration-related articles published between 2013 and 2023. Furthermore, we provide an overview of the evolving trends in the literature on these topics. The findings show a significant increase in research attention to multi-omics and data integration, particularly in recent years, and the growing interest suggests that research in these areas will continue to gain attention because integrating multi-omics data is critical for understanding the mechanisms underlying health, disease, and other physiological conditions. This could lead to developing new methodologies algorithms and or validating existing approaches across various research settings.

REFERENCES

1. Zhong Y, Peng Y, Lin Y, et al. MODILM: towards better complex diseases classification using a novel multi-omics data integration learning model. *BMC Med Inform Decis Mak.*2023;23(82). doi:10.1186/s12911-023-02173-9.

2. Gallo Cantafio ME, Grillone K, Caracciolo D, et al. From Single Level Analysis to Multi-Omics Integrative Approaches: A Powerful Strategy towards the Precision Oncology. *High-throughput.* 2018;7(4):33. doi:10.3390/ht7040033.

3. Ritchie MD, Holzinger ER, Li R, Pendergrass SA, Kim D. Methods of integrating data to uncover genotype-phenotype interactions. *Nat Rev Genet.* 2015;16(2):85-97. doi:10.1038/nrg3868.

4. Subramanian I, Verma S, Kumar S, Jere A, Anamika K. Multi-omics Data Integration, Interpretation, and Its Application. *Bioinform Biol Insights.* 2020;14:1177932219899051. doi:10.1177/1177932219899051.

5. Cavill R, Jennen D, Kleinjans J, Briedé JJ. Transcriptomic and metabolomic data integration. *Brief Bioinform.*2016;17(5):891-901. doi:10.1093/bib/bbv090.

6. Lin E, Lane HY. Machine learning and systems genomics approaches for multi-omics data. *Biomark Res.* 2017;5:2. doi:10.1186/s40364-017-0082-y.

7. Ivanisevic T, Sewduth RN. Multi-Omics Integration for the Design of Novel Therapies and the Identification of Novel Biomarkers. *Proteomes.* 2023:11(4):34. doi:10.3390/proteomes11040034.

8. van Eck NJ, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics.* 2010;84(2):523-538. doi:10.1007/s11192-009-0146-3.

9. Chen C. Cite Space II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *J Am Soc Inf Sci Technol.* 2006;57(3):359-377. doi:10.1002/asi.20317.

10. Donthu N, Kumar S, Mukherjee D, Pandey N, Lim WM. How to conduct a bibliometric analysis: An overview and guidelines. *J Bus Res.* 2021;133:285-296.

11. Hashem E AR, Md Salleh NZ, Abdullah M, et al. Research trends, developments, and future perspectives in brand attitude: A bibliometric analysis utilizing the Scopus database (1944-2021). *Heliyon.* 2022;9(1):e12765. doi:10.1016/j.heliyon.2022.e12765.

12. Aria M, Cuccurullo C. Bibliometrix: An R-tool for Comprehensive Science Mapping Analysis. *J Informetr.* 2017;11:959-975. doi:10.1016/j.joi.2017.08.007.

13. Argelaguet R, Velten B, Arnol D, et al. Multi-Omics Factor Analysis-a framework for unsupervised integration of multi-omics data sets. *Mol Syst Biol.* 2018;14(6):e8124. doi:10.15252/msb.20178124.

14. Chaudhary K, Poirion OB, Lu L,Garmire LX. Deep Learning-Based Multi-Omics Integration Robustly Predicts Survival in Liver Cancer. *Clin cancer res.* 2018;24(6):1248-1259. doi:10.1158/1078-0432.CCR-17-0853.

15. Pang Z, Zhou G, Ewald J, et al. Using MetaboAnalyst 5.0 for LC-HRMS spectra processing, multi-omics integration and covariate adjustment of global metabolomics data. *Nat Protoc.* 2022;17(8):1735-1761. doi:10.1038/s41596-022-00710-w.

16. Pinu FR, Beale DJ, Paten AM, et al. Systems Biology and Multi-Omics Integration: Viewpoints from the Metabolomics Research Community. *Metabolites.* 2019;9(4):76. doi:10.3390/metabo9040076.

17. Mertins P, Mani DR, Ruggles KV, et al. Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature.* 2016;534(7605):55-62. doi:10.1038/nature18003.

18. Wang TJ, Larson MG, Vasan RS. Metabolite profiles and the risk of developing diabetes. *Nature Medicine.* 2011;17(4):448-453. doi:10.1038/nm.2307