Original

Open access and data sharing in cancer stem cells research

Rut Lucas-Domínguez[a,b,c], Antonio Vidal-Infer[a,b,*], Adolfo Alonso-Arroyo[a,b],
Beatriz Tarazona-Álvarez[d], Máxima Bolaños-Pizarro[a,e], Vanessa Paredes-Gallardo[d]

[a] Department of History of Science and Information Science, University of Valencia, Valencia, Spain
[b] UISYS, Joint Research Unit, University of Valencia, Valencia, Spain. Associated Unit Research Institute for Higher Education and Science (INAECU) UC3M-UAM, Spain
[c] CIBERONC, Valencia, Spain
[d] Department of Stomatology, School of Medicine and Dentistry, University of Valencia, Valencia, Spain
[e] Research Institute of Higher Education and Science (INAECU) UC3M-UAM, Spain

ARTICLE INFO

ABSTRACT

Objective: Understanding and targeting cancer stem cells (CSC) are areas of active research in oncology and further dissemination of scientific results is urgently needed to accelerate the comprehension of their involvement in tumor heterogeneity, resistance to conventional therapies and metastasis. In the context of open science, open access and data sharing, the aim of this study is to assess the current practices of cancer researchers in terms of publications and dissemination of research data.
Method: A bibliometric study was conducted based on a bibliographic search using the Web of Science, and CSC articles with at least one Spanish affiliation were retrieved. A comparative study of the openness criteria of the journals was carried out, as well as an analysis of the associated data attached to the articles as supplementary material or deposited in repositories.
Results: 708 documents from 282 journals were retrieved. 303 articles contained associated research data, mostly published in Q1 journals, demonstrating a correlation between impact journals and their commitment to quality improvement. Supplementary material was the preferred method of data sharing, being pdf the most used file type. Only 69 articles mentioned datasets deposited in repositories, mainly from genomic nature. The main cancers studied were digestive and gastrointestinal, neurological, breast, hematological and respiratory, which coincided with the tumor types with the highest incidence and mortality, and in which the presence of CSC has been described.
Conclusions: Researchers and publishers have become more aware of open science practices, but there is still a need for data quality according to the FAIR principles.

© 2024 SESPAS. Published by Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Acceso abierto y uso compartido de datos de investigación sobre células madre cancerosas

RESUMEN

Objetivo: La investigación activa en células madre cancerosas (CMC) requiere la difusión de los resultados científicos para acelerar la comprensión de la heterogeneidad tumoral, la resistencia a las terapias convencionales y la metástasis. En el contexto de la ciencia abierta, el acceso abierto y el intercambio de datos, el objetivo de este estudio es evaluar las prácticas actuales de los investigadores del cáncer en términos de publicaciones y difusión de datos de investigación.
Método: Se realizó un estudio bibliométrico a partir de una búsqueda bibliográfica utilizando Web of Science y se recuperaron artículos sobre CMC firmados por al menos una institución española. Se realizó un análisis de los criterios de apertura de las revistas, así como de los datos asociados a los artículos.
Resultados: Se recuperaron 708 documentos de 282 revistas. De ellos, 303 artículos contenían datos de investigación asociados, mayoritariamente publicados en revistas del primer cuartil, mostrando una correlación entre las revistas de impacto y su calidad. El material suplementario fue el método preferido para compartir datos, siendo el *pdf* el formato más utilizado. Solo 69 artículos mencionaron conjuntos de datos depositados en repositorios, principalmente de naturaleza genómica. Los cánceres estudiados fueron el digestivo y gastrointestinal, el neurológico, el de mama, el hematológico y el respiratorio, coincidiendo con los tumores con mayor incidencia y mortalidad, y en los que se ha descrito la presencia de CMC.
Conclusiones: Los investigadores y los editores se han vuelto más conscientes de las prácticas de ciencia abierta, pero todavía existe la necesidad de implementar la calidad de los datos de acuerdo con los principios FAIR.

© 2024 SESPAS. Publicado por Elsevier España, S.L.U. Este es un artículo Open Access bajo la CC BY-NC-ND licencia (http://creativecommons.org/licencias/by-nc-nd/4.0/).

* Corresponding author.
E-mail address: Antonio.Vidal-Infer@uv.es (A. Vidal-Infer).

Introduction

Cancer remains the second leading cause of death worldwide, with almost 10 million deaths in 2022 and nearly 20 million new cases, and an estimated 35 million new cases by 2050, making it a major public health concern. It is predicted that about one in five men or women will develop cancer in their lifetime, while about one in nine men and one in 12 women will die from cancer.[1,2] Overall, lung and colorectal cancer are the main causes of death, followed by liver, breast and stomach cancer. The higher incidence is mostly due to lung cancer, breast cancer, colorectal, prostate and stomach cancer.[3]

The treatment of cancer patients depends on the histology and tumor stage. Surgical resection is the preferred treatment for patients in the early stages of the disease, while chemotherapy and radiotherapy are the standard treatments for those diagnosed in advanced stages. However, even with the development of less toxic and more effective chemotherapy, the lack of precise therapeutic targets and continued tumor resistance to treatments has led to the emergence of precision oncology. This approach is based on new therapies that aim to eradicate oncogenic molecular alterations, select patients, and administer drugs to malignancies with these alterations.[4] More recently, immunotherapy has become the most attractive treatment option for various types of cancer with very encouraging results, including melanoma and lung cancer, among others.[5] Nevertheless, despite the considerable advances that have been made in treatment modalities, many cancer patients experience recurrence and metastasis.[6] In this sense, it is known that cancer cells are heterogeneous and there is solid evidence indicating that chemoresistance, tumor progression and metastasis are linked to a subpopulation of cells with "stem" characteristics called Cancer Stem Cells (CSC).[4,7] Understanding the mechanisms that regulate the properties of CSC and developing therapeutic strategies specifically targeting these cells are areas of active research, aimed at improving treatment outcomes and survival rates for cancer patients.[7,8] However, few studies have attempted to analyse the scientific production on CSC in terms of derived articles, the content of which is limited only to focus on one or a few specific cancer types.[9] Furthermore, the extent of the raw data produced by CSC research is also not known.

In this respect, the availability of open access (OA) publications and the accessibility to the high quality research datasets that underpin these articles is crucial, as it can improve the dynamics of cancer research.[10] Among the main problems faced by researchers in the field of oncology, it is worth highlighting the limitations of patient samples obtained by invasive techniques, the lack of reproducibility of published results and the need for OA articles; and even more complex, finding the datasets that support these results.[11] Research raw data may be deposited in repositories or included as supplementary material with the papers. The benefit of data sharing has been described in the acceleration of research in medicine[12] and recently during the last pandemic, by SARS-CoV-2.[13,14]

The objective of this work is to evaluate the scientific production on CSC, with particular emphasis on the study of current practices related to OA publications and research data sharing, to examine the availability and type of raw data sets released through scientific publications, signed by at least one Spanish affiliation.

Method

A bibliometric study was conducted in the following stages:

1) Search strategy. The terms of the search equation were identified from the classification of cancer types established by the U.S. National Cancer Institute (NCI)[15] and combined with the terms "Cancer* Stem Cell*" OR "Cell* Cancer* Stem" OR "Stem Cell* Cancer*" (see Supplementary file 1). The bibliographic search was carried out through the Science Citation Index Expanded (SCIE) of the Web of Science (WoS) in January 2021, and the selected period was until December 31, 2020, limiting the results to the Country: Spain. An Access database was generated with a total of 708 retrieved documents which contained the following information: article title, authors' names, institutional affiliations, journal title, financial support, keywords, and data availability statement.

2) Comparative study of scientific publications in CSC was developed analysing the journal's openness criteria classified into quartiles (Q) based on the 2019 SCIE edition of the Journal Citation Reports of WoS. Additional data were collected for each journal: publisher, journal's access modality (gold OA, hybrid, traditional), possibility of storing manuscripts in thematic or institutional repositories, reuse policy, policy regarding publication on the official website or by the author, statement of policy on supplementary material, and statement of policy related to deposit of research data.
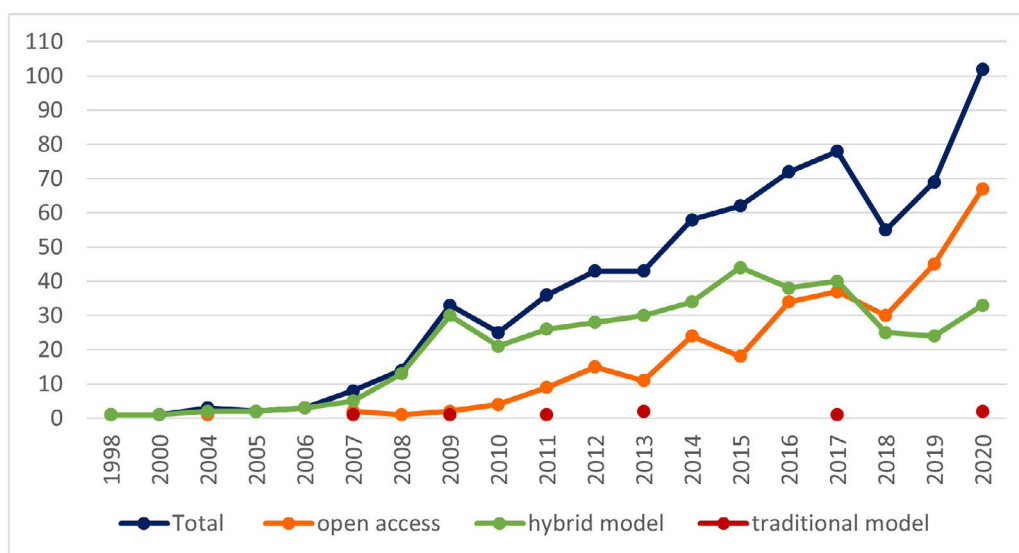

Figure 1. Temporal evolution of scientific production on cancer stem cells (blue line), classified as papers derived from open access (orange line), hybrid (green line) and traditional model (red line) journals.

3) Analysis of research data attached to the article through dissemination as supplementary material or deposit of research data in repositories. To identify publications with research data available in the form of supplementary material, a secondary search was performed by two independent researchers reviewing manually each paper. The number and types of files located on the articles as supplemental material were registered even if a single article included several different files. On the other hand, the searches of the publications with raw research data deposited in repositories were executed in WoS-SCIE combining the initial search equation related to CSC with the terms referring to the main thematic repositories of biomedical sciences and cancer.[13]

4) The title, abstract and keywords of the articles were examined in order to perform a thematic study of the papers on CSC, depending on the type of cancer investigated and the anatomical location according to the U.S. NCI classification.[15]

5) Descriptive analyses of variables and crosstables were performed using Microsoft Access database and Excel software. Authors, institutional affiliations, and funding entities were normalised manually. Analysis and visualisation of networks of authors and institutions were performed using Pajek software.

Results

Analysis of scientific publications in the field of CSC

Scientific production in CSC resulted in 708 documents from 282 journals, signed by at least one Spanish institution, being 507 articles and 201 reviews (Fig. 1). Analyses of publishers' OA policies showed that 400 articles came from 197 hybrid journals, 300 articles came from 77 gold OA journals and 8 articles were published in 8 traditional journals (Fig. 1). The distribution of publications was 457 documents in Q1 journals, 163 in Q2 journals, 66 in Q3 journals and 22 in Q4 journals. The most productive journals in CSC research with the highest proportion of associated research data were Scientific Reports and PLoS One (see Supplementary Table 1).

Regarding OA policies, Table 1 shows that, of the five variables analyzed for all 282 journals, the great majority (except for 5-10 journals) allowed the deposit of articles in thematic or institutional repositories, as well as their re-use and publication on the website, either directly (around 30%) or under certain conditions (around 70%). In terms of associated research data to the articles, 90% of journals accepted the inclusion of supplementary material unconditionally, and 60% suggested depositing data in repositories, while 23% made it mandatory.

Networks of authors and institutions

Scientific production on CSC was signed by 4,001 authors from 828 institutions, identifying 249 Spanish affiliations. The number of authors who participated in a single article were 3,153 (78.8%), which contrasted with the fact that 27 authors signed more than 10 papers and 7 authors exceeded 20 published articles. In parallel, 7 institutions appeared in more than 50 papers. The most productive authors and institutions related to CSC research signed by at least one Spanish affiliation are described in Supplementary Table 2.

Figure 2 A shows the 91 authors who are part of collaborative networks with more than 5 publications in common. There are 17 different networks, ranging from those formed by only two members to the most participative network with 16 different authors. The most prolific author is Javier A Menéndez, with 38 published papers and he leads the network with the most representatives, Begoña Martín Castillo, and Cristina Oliveras Ferrarós are his closest collaborators with 24 and 18 papers in common.

Table 1
Analysis of variables concerning the availability of articles and raw data of the 282 journals, distributed by quartiles.

Q	Storage in thematic or institutional repositories				Reuse				Publication in website				Statement of supplementary material				Statement of deposited research data		
	A	NA	AC	NS	A	NA	AC	NS	A	NA	AC	NS	A	NA	AC	NS	M	S	NS
1	42 (52.5%)	1 (50.0%)	119 (62.0%)	2 (25.0%)	42 (54.5%)	1 (100.0%)	118 (60.5%)	3 (33.3%)	46 (61.3%)	1 (100.0%)	114 (57.6%)	3 (37.5%)	155 (60.8%)	2 (66.7%)	3 (50.0%)	4 (22.2%)	45 (68.2%)	98 (57.6%)	21 (45.7%)
2	19 (23.8%)	0 (0.0%)	49 (25.5%)	0 (0.0%)	18 (23.4%)	0 (0.0%)	50 (25.6%)	0 (0.0%)	19 (25.3%)	0 (0.0%)	49 (24.7%)	0 (0.0%)	61 (23.9%)	0 (0.0%)	3 (50.0%)	4 (22.2%)	11 (16.7%)	49 (28.8%)	8 (17.4%)
3	13 (16.3%)	1 (50.0%)	17 (8.9%)	1 (12.5%)	11 (14.3)	0 (0.0%)	20 (10.3%)	1 (11.1%)	6 (8.0%)	0 (0.0%)	25 (12.6%)	1 (12.5%)	29 (11.4%)	0 (0.0%)	0 (0.0%)	3 (16.7%)	8 (12.1%)	16 (9.4%)	8 (17.4%)
4	6 (7.5%)	0 (0.0%)	7 (3.6%)	5 (62.5%)	6 (7.8%)	0 (0.0%)	7 (3.6%)	5 (55.6%)	4 (5.3%)	0 (0.0%)	10 (5.1%)	4 (50.0%)	10 (3.9%)	1 (33.3%)	0 (0.0%)	7 (38.9%)	2 (3.0%)	7 (4.1%)	9 (19.6%)
T	80 (28.4%)	2 (0.7%)	192 (68.1%)	8 (2.8%)	77 (27.3%)	1 (0.4%)	195 (69.1%)	9 (3.2%)	75 (26.6%)	1 (0.4%)	198 (70.2%)	4 (2.8%)	255 (90.4%)	3 (1.1%)	6 (2.1%)	18 (6.4%)	66 (23.4%)	170 (60.3%)	46 (16.3%)

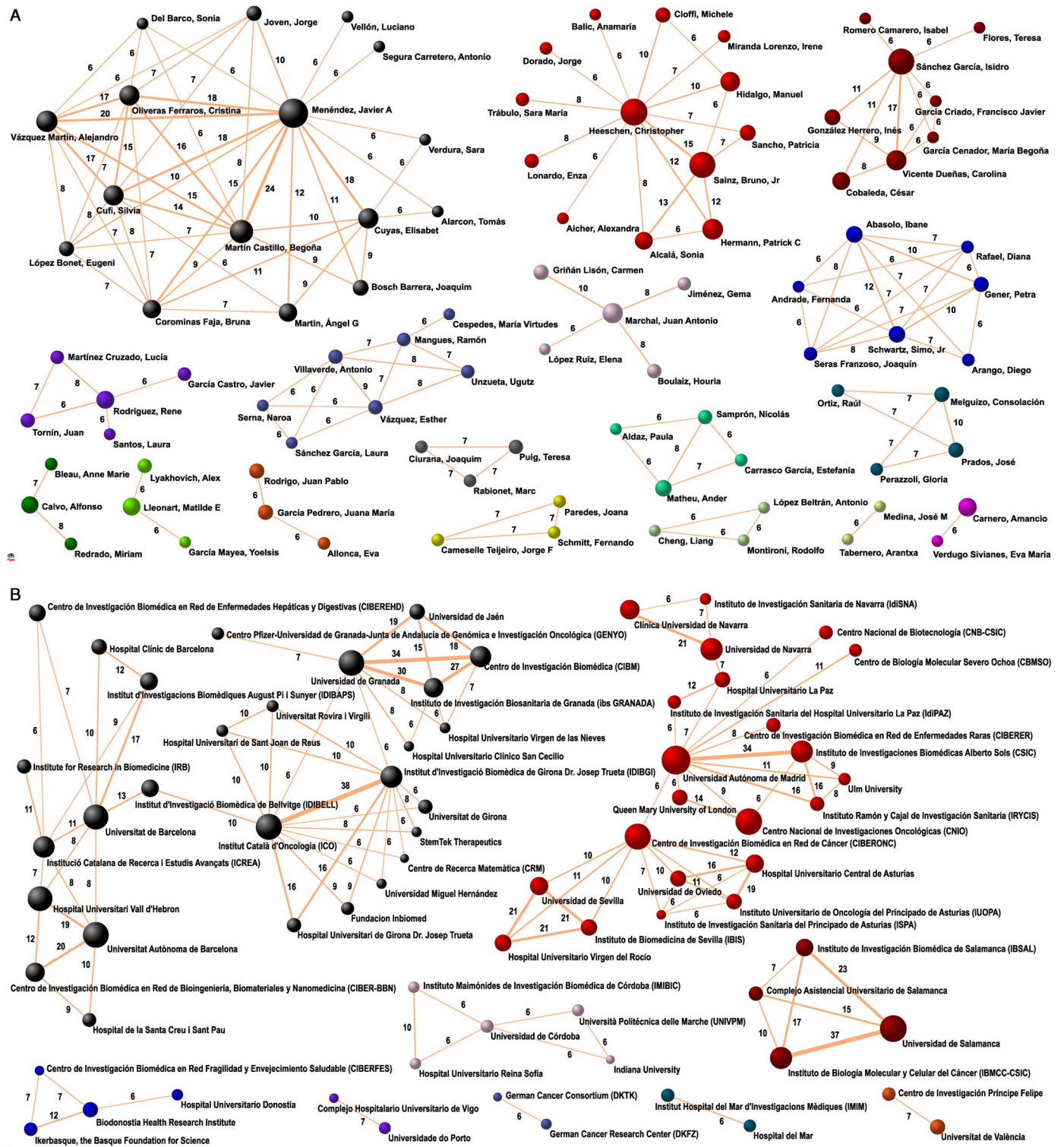A: allowed; NA: not allowed; NS: not specified; M: mandatory; Q: quartil; S: Ssggested.

Figure 2. Networks of authors (**A**) and institutions (**B**) involved in the publications of articles related to cancer stem cells signed by at least one Spanish affiliation. The thickness of the spheres represents the number of articles published in the study.

The network of institutions represents the 71 institutions that collaborated with another institution on more than 5 papers. Figure 2 B shows 9 different networks, highlighting two large networks with 28 and 22 different institutions. Institut Català d'Oncologia (ICO) is the main network with 28 institutions, which works closely with Institut d'Investigació Biomèdica de Bellvitge (IDIBELL), Institut d'Investigació Biomèdica de Girona Dr. Josep Trueta (IDIBGI) and the University of Granada. This network brings

together a larger number of institutions through IDIBELL, which acts as an intermediary node between the ICO and the University of Barcelona. All the affiliations involved in the network are Spanish. The second network of 22 different institutions is led by the Universidad Autónoma de Madrid, which establishes links with up to 11 institutions, with Centro de Investigación Biomédica en Red en Oncología (CIBERONC) and the Hospital Universitario la Paz acting as intermediaries to connect with institutions in Seville, Asturias

Table 2
Mentions of data availability statement found in the 708 CSC articles analysed by journal quartiles and financial support.

Data availability statement	Q1 n (%)	Q2 n (%)	Q3 n (%)	Q4 n (%)	Total n	Funding n
None	190 (41.6)	100 (21.9)	46 (10.1)	19 (4.2)	355	293
Supplementary material	206 (45.1)	46 (10.1)	10 (2.2)	2 (0.4)	264	254
Repository	3 (0.7)	1 (0.2)	1 (0.2)	0	5	4
Supplem plus Repository	53 (11.6)	10 (2.2)	1 (0.2)	0	64	63
On request to the author	0	0	1 (0.2)	0	1	1
Available in the article	3 (0.7)	3 (0.7)	1 (0.2)	0	7	6
Others	2 (0.4)	3 (0.7)	6 (1.3)	1 (0.2)	12	10
Total per quartil	457	163	66	22	708	631

and Navarra. On this occasion, Universidad Autónoma de Madrid collaborated with the foreign universities of Ulm (11 papers) and Queen Mary University of London (6 papers).

Research data on CSC through dissemination as supplementary material and deposited in repositories

The evaluation of the mode of diffusion of the research data on CSC in terms of data availability statement (Table 2) indicated that 37.3% of the articles contained data as supplementary material associated with the published article (264 documents), 0.7% were publications that included the research data deposited in repositories (5 documents) and 9.0% of the articles on CSC (64 documents) contained both options for the availability of research data.

The relationship between the number of CSC articles retrieved with research data and the journal impact factor based on their positioning showed that 57.3% of the articles published in Q1 journals contained associated data (45.1% as supplementary material and 11.6% both supplementary and deposited in repositories), while only 12.5% of papers published in Q2 journals reported complementary research data. Regarding to the CSC papers produced with financial support (89%), the percentage of funded articles including associated data represented the 50.9% (40.3% in supplementary, 1.6% deposited in repositories, and 10% in supplementary plus repository) (Table 2).

Figure 3 A represents the analysis of supplementary material associated with the published articles, which was derived mainly from Q1 publications. The results showed that .pdf format was the most common file type (35.3%), followed mainly by .jpeg or .tiff image files (20.3%), .xls files (14.0%), .doc or .text files (12.8%) and video files (5.2%). On the other hand, in terms of the analysis and classification of the main repositories used by researchers, the datasets content was mainly of a genetic nature, with the Gene Expression Omnibus (U.S. National Center for Biotechnology Information) being the most prominent, containing a total of 50 datasets (64.1%), followed by the Array Express Archive of Functional Genomics Data which contained 9 datasets (11.5%) (Fig. 3 B).

Overview of the most common cancers studied in relation to CSC

The analysis of the title, abstract and keywords of the 708 articles according to the type of cancer studied showed that 28.7% were related to basic research on CSC and contained less associated research data in comparison with the most frequently studied typologies (breast cancer, brain cancer, pancreatic cancer, leukemia, liver cancer, colorectal cancer, and lung cancer) (Fig. 4 A). These results are similar regarding to the anatomical location affected by the neoplasms (Fig. 4 B).

Discussion

CSC have the ability to self-renew and differentiate into various cell types within the tumor mass, contributing to tumor heterogeneity; are thought to be responsible for initiating tumor formation, which is often resistant to conventional cancer therapies; and have been implicated in the process of metastasis, as they can migrate from the primary tumor site to distant organs and initiate secondary tumor growth.[8] Therefore, understanding and targeting CSC are areas of active research in oncology, making the subsequent dissemination of scientific results urgent to accelerate the knowledge of cancer, which is a major public health problem and in which a large amount of economic resources is invested.[16] This study shows that the international scientific production in the CSC resulted in 708 documents signed by at least one Spanish institution, with the first article in 1998 and a 30-fold increase in 2009, reaching 100 articles in 2020. A decrease in publications is observed in 2018, which coincides with the discontinuation of the journal *Oncotarget* from the WoS database. From 2019, scientific production recovers and the journals of the MDPI publishing group begin to gain weight, although it is the Nature publishing group that leads the ranking of publications with associated research data. CSC articles showed the participation of 4,001 authors from 828 institutions, of which only 249 affiliations were Spanish, highlighting 7 authors with more than 20 published articles and 7 institutions that appeared in more than 50 papers, being the top producers the Autonomous University of Madrid, the University of Salamanca, Centro Nacional de Investigaciones Oncológicas (CNIO), ICO, CIBERONC and the Universidad Autónoma de Madrid.

Analyses of 288 journals openness criteria demonstrates that over the years there has been an augment in publications in gold OA journals compared to the hybrid model, surpassing it as of 2018, with hardly any publications in commonly known traditional journals. Furthermore, around 96% of journals publishing CSC articles allowed directly or with some conditions the storage of manuscripts in thematic or institutional repositories, reuse, and publication in websites. This behaviour is particularly consistent with the promotion of OA statements in scientific research between 2002 and 2007 and the subsequent mandatory OA publication of publicly funded research.[17,18]

It is in this context that the concept of Open Science (OS) is framed, meaning unrestricted access to all aspects of research, so that everyone can follow, use and participate in science. OS includes a growing list of other open areas of interests, not only OA to the publications but also open research data, which is a valuable resource that is often expensive and time-consuming to generate.[19] In fact, one of the practices that has gained great momentum in recent years because of its ability to improve quality of research is the sharing of research data.[20,21] To promote data sharing in cancer research, initiatives such as OA journals, data repositories, and collaborative research networks have been established. For example, The Cancer Genome Atlas (TGCA) is a project launched in 2005 by the NCI and the National Human Genome Research Institute of the USA to accelerate understanding of the molecular basis of cancer by applying genomic analysis and protein expression profiling maximizing data sharing. In 2013, the Global Alliance for
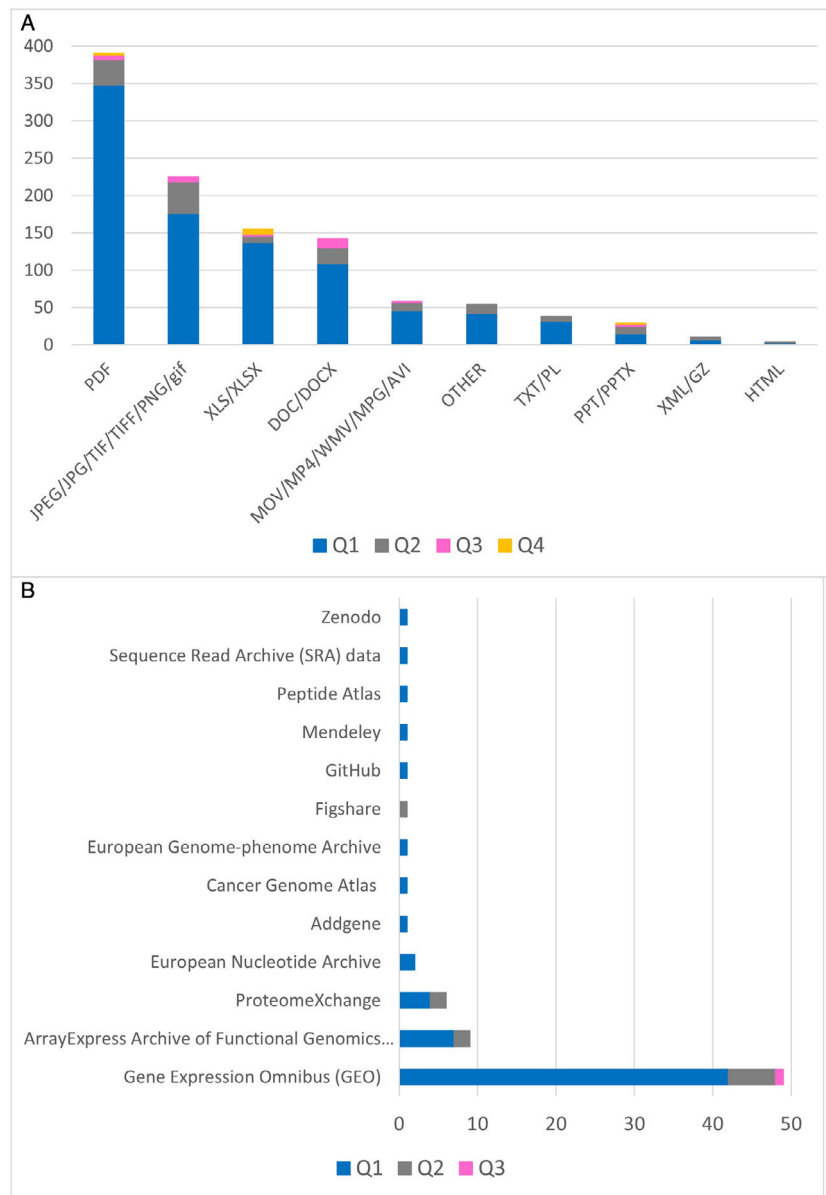
Figure 3. Analysis of the research data associated with the articles classified by quartile of journals, in the form of number and type of files attached as supplementary material **(A)** and datasets deposited in different repositories **(B)**.

Genomics (GA4GH) was established as a common framework to enable responsible, voluntary and secure sharing of clinical and genomic data.[22] In Europe, the Oncology Data Network (ODN) for all cancer types was developed in 2017 with accessibility to all cancer centers and patients.[23] Its objective is to establish a comprehensive infrastructure of real-world information on cancer care. The final aim is to inform precision medicine and validate specific therapeutic approaches for each cancer subtype by sharing data derived from routine clinical experience.[24]

As a further step, the rationale for this study was therefore based on the availability of CSC research and datasets shared as supplementary information in articles or deposited in repositories. Our results show that 303 out of 708 articles contained associated research data (47%), mostly published in Q1 journals (57.3%), demonstrating a relationship between impact journals and their commitment to improving the quality, transparency and reproducibility of the communicated research.

Supplementary material was the preferred method of data sharing, with pdf format being the most common file type used,

followed by image and xls files. These findings suggest a similar behavior to previous studies of data sharing in regenerative medicine and stem cell research (45.9%).[25] Furthermore, these are promising results, although not sufficient, when compared with the availability of associated data in other fields such as emergency medicine (9.4%),[26] general and internal medicine (9.5%),[27] dentistry (7.6%) or substance abuse (4.7%).[28] In contrast, only 69 articles mentioned data sets deposited in repositories, showing that this is still an uncommon practice, closely related to omics research, referring to data from genomic, transcriptomic, proteomic and metabolomic analyses, and which ranks first in the GEO repository. These results highlight the need to ensure that the research data generated is in line with the FAIR (Findable, Accessible, Interoperable and Reusable) principles developed to identify good practice in data sharing.[10]

Regarding the main organs affected by cancer in the 708 papers studied, first position is digestive/gastrointestinal cancer (17.2%), followed by neurologic cancer (14.2%), breast cancer (13.5%), hematologic/blood cancer (8.7%) and respiratory/thoracic cancer (3.2%).
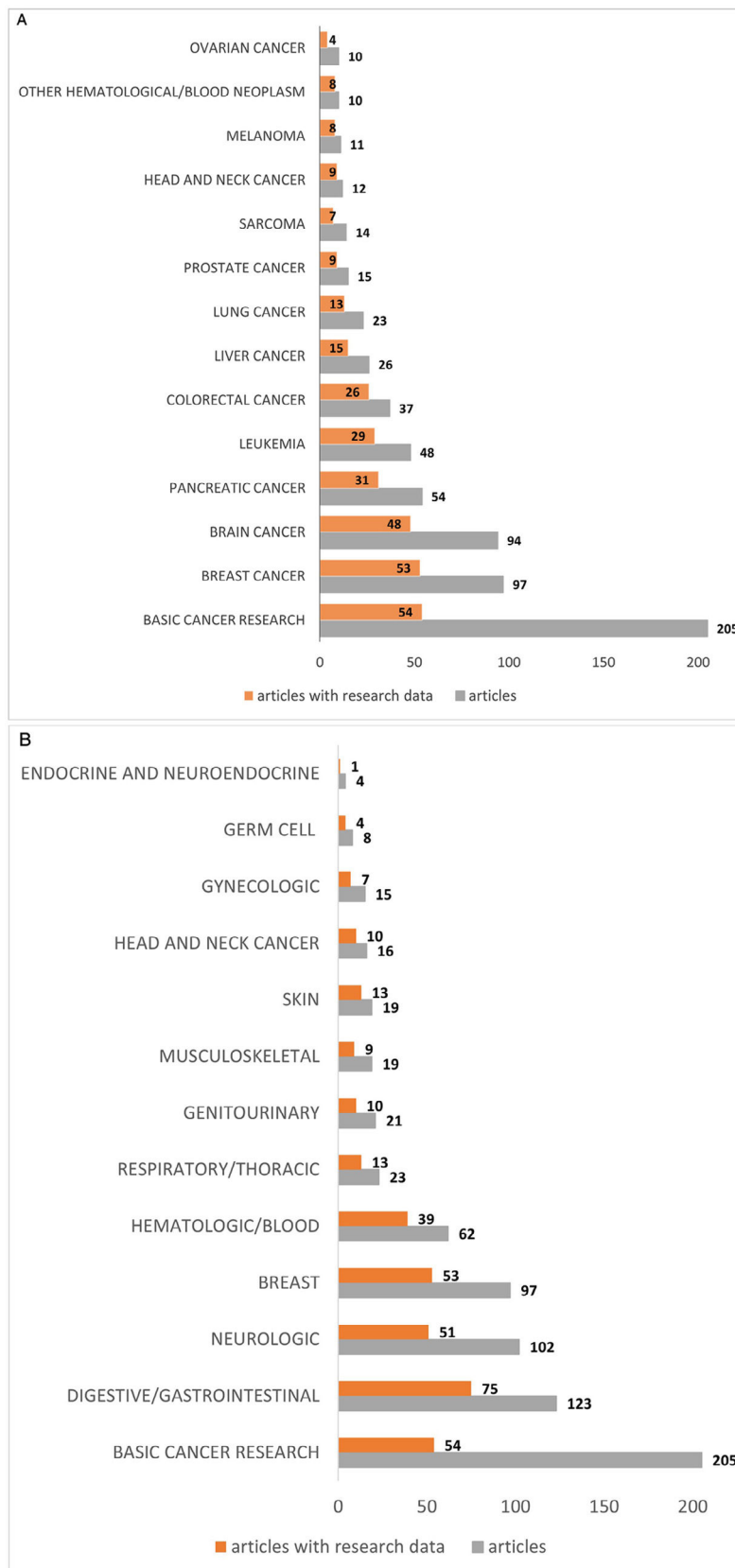
Figure 4. Thematic analysis of cancer stem cells-related publications and the presence of associated research data, classified by cancer typology **(A)** and body location **(B)**.

These results are consistent with the most commonly diagnosed cancers and the leading causes of death,[1,3] but also indicate those cancers where the presence of CSC has been identified as play-ing a significant role in the disease. CSC were first identified in leukemia and are thought to contribute to treatment resistance and disease progression.[29] Glioblastoma is a highly aggressive

type of brain cancer in which CSCs have the ability to resist radiotherapy and chemotherapy, contributing to tumor recurrence after treatment.[9] Similarly, in colorectal cancer, CSC play an important role in tumor initiation and progression, as well as treatment resistance and recurrence, and in breast cancer, CSC are able to self-renew and generate daughter cells that make up the tumor, contributing to its heterogeneity and ability to resist treatment.[30]

Our findings in cancer research through the CSC study show that OA and data sharing policies are being implemented, but at a much slower rate than might be expected, and these data are consistent with other studies on the availability and capture of study protocols in oncology.[31] Journals and publishers have a key role to play in promoting OS policies, and researchers should adopt guidelines accepted by the scientific community when archiving their research results.[32] In addition, biomedical research institutions have the responsibility to develop data management and sharing plans supervising OS practices,[33] and government agencies should recognise researchers and institutions that adhere to OS mandates.[34]

Limitations

This study has analysed articles on CSC indexed in WoS-SCIE, so it is possible that other articles from journals indexed in other databases have not been included. The fieldwork was performed in January 2021, thus subsequent publications were not included. The linguistic bias of WoS might underrepresent publications in languages other than English.

Availability of databases and material for replication

The data generated and used during this research are openly available from Zenodo.org public repository at https://doi.org/10.5281/zenodo.11370781.

What is known about the topic?

Understanding cancer stem cells is an area of active research and communication of scientific results is needed. In the context of open science, open access and data sharing, the aim of this study is to assess the current practices of cancer researchers in relation to the dissemination of scientific data.

What does this study add to the literature?

Less than half of the cancer stem cell articles contained research data, and most were published in Q1 and Q2 journals, demonstrating a correlation between impact journals and their commitment to quality improvement.

What are the implications of the results?

Researchers and publishers have become more aware of the need to publish in open access journals and to deposit data in recent years, but there is still a need to raise awareness of data quality according to the FAIR principles.

Editor in charge

Cristina Candal Pedreira

Transparency declaration

The corresponding author, on behalf of the other authors, guarantee the accuracy, transparency and honesty of the data and information contained in the study, that no relevant information has been omitted and that all discrepancies between authors have been adequately resolved and described.

Authorship contributions

R. Lucas-Domínguez and A. Vidal-Infer: conceptualization, data extraction, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, visualization, writing original draft, review and editing. A. Alonso-Arroyo: data curation, formal analysis, investigation, methodology, visualization, writing original draft, review and editing. B. Tarazona-Álvarez, M. Bolaños-Pizarro and V. Paredes-Gallardo: data extraction, formal analysis, investigation, methodology, review and editing.

Funding

Conflicts of interest

None.

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version available at https://doi.org/10.1016/j.gaceta.2024.102432.

References

1. Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024;74:229–63.
2. Ferlay J, Ervik M, Lam F, et al. Global Cancer Observatory: Cancer Today. Lyon, France: International Agency for Research on Cancer; 2024. (Accessed 11 April 2024). Available at: https://gco.iarc.who.int/today/.
3. World Health Organization. Global cancer burden growing, amidst mounting need for services. WHO; 2024. (Accessed 30 April 2024). Available at: https://www.who.int/news/item/01-02-2024-global-cancer-burden-growing–amidst-mounting-need-for-services.
4. Lawson DA, Kessenbrock K, Davis RT, et al. Tumour heterogeneity and metastasis at single-cell resolution. Nat Cell Biol. 2018;20:1349–60.
5. Baik CS, Rubin EH, Forde PM, et al. Immuno-oncology clinical trial design: limitations, challenges, and opportunities. Clin Cancer Res. 2017;23:4992–5002.
6. Ganesh K, Massague J. Targeting metastatic cancer. Nat Med. 2021;27:34–44.
7. Yang L, Shi P, Zhao G, et al. Targeting cancer stem cell pathways for cancer therapy. Signal Transduct Target Ther. 2020;5:8.
8. Batlle E, Clevers H. Cancer stem cells revisited. Nat Med. 2017;23:1124–34.
9. Song S, Wu H, Wang F, et al. Global research trends and hotspots on glioma stem cells. Front Oncol. 2022;12:926025.
10. Vesteghem C, Brøndum RF, Sønderkær M, et al. Implementing the FAIR data principles in precision oncology: review of supporting initiatives. Brief Bioinform. 2020;21:936–45.
11. Rodgers P, Collings A. What have we learned? Elife. 2021;10:e75830.
12. Kadakia KT, Beckman AL, Ross JS, et al. Leveraging open science to accelerate research. N Engl J Med. 2021;384:e61.
13. Lucas-Dominguez R, Alonso-Arroyo A, Vidal-Infer A, et al. The sharing of research data facing the COVID-19 pandemic. Scientometrics. 2021:1–16.
14. Moorthy V, Henao Restrepo AM, Preziosi M-P, et al. Data sharing for novel coronavirus (COVID-19). Bull World Health Organ. 2020;98:150.
15. National Cancer Institute. A to Z list of cancer types - NCI. 2024. (Accessed 10 May 2024). Available at: https://www.cancer.gov/types.
16. Chen S, Cao Z, Prettner K, et al. Estimates and projections of the global economic cost of 29 cancers in 204 countries and territories from 2020 to 2050. JAMA Oncol. 2023;9:465–72.

17. BOAI20. The Budapest Open Access Initiative: 20th Anniversary Recommendations. 2022. (Accessed 16 May 2024). Available at: https://www.budapestopenaccessinitiative.org/boai20/.
18. Björk B-C. Open access to scientific articles: a review of benefits and challenges. Intern Emerg Med. 2017;12:247–53.
19. UNESCO. Recommendation on Open Science. (Accessed 15 May 2024). Available at: https://www.unesco.org/en/open-science/about.
20. Munafò MR, Nosek BA, Bishop DVM, et al. A manifesto for reproducible science. Nat Hum Behav. 2017;1:0021.
21. World Health Organization. New WHO policy requires sharing of all research data. WHO; 2022. (Accessed 19 May 2024). Available at: https://www.who.int/news/item/16-09-2022-new-who-policy-requires-sharing-of-all-research-data.
22. Siu LL, Lawler M, Haussler D, et al. Facilitating a culture of responsible and effective sharing of cancer genome data. Nat Med. 2016;22:464–71.
23. Kerr D, Arnold D, Blay J-Y, et al. The Oncology Data Network (ODN): a collaborative European data-sharing platform to inform cancer care. Oncologist. 2020;25:E1–4.
24. Woolmore A, Arnold D, Blay J-Y, et al. The Oncology Data Network (ODN): methodology, challenges, and achievements. Oncologist. 2020;25:e1428–32.
25. Aleixandre-Benavent R, Lucas-Domínguez R, Sixto-Costoya A, et al. The sharing of research data in the cell & tissue engineering area: is it a common practice? Stem Cells Dev. 2018;27:717–22.
26. Sixto-Costoya A, Aleixandre-Benavent R, Lucas-Dominguez R, et al. The emergency medicine facing the challenge of open science. Data. 2020;5:28.
27. Tarazona-Alvarez B, Zamora-Martinez N, Garcia-Sanz V, et al. Public availability of research data in general and internal medicine journals. Zenodo; 2021. (Accessed 18 November 2021). Available at: https://zenodo.org/record/5709924.
28. Vidal-Infer A, Aleixandre-Benavent R, Lucas-Domínguez R, et al. The availability of raw data in substance abuse scientific journals. Journal of Substance Use. 2018;0:1–5.
29. Bonnet D, Dick JE. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. Nat Med. 1997;3:730–7.
30. Feng Y, Spezia M, Huang S, et al. Breast cancer development and progression: risk factors, cancer stem cells, signaling pathways, genomics, and molecular pathogenesis. Genes & Diseases. 2018;5:77–106.
31. Collins GS, Whittle R, Bullock GS, et al. Open science practices need substantial improvement in prognostic model studies in oncology using machine learning. J Clin Epidemiol. 2024;165:111199.
32. Hamilton DG, Page MJ, Finch S, et al. How often do cancer researchers make their data and code available and what factors are associated with sharing? BMC Med. 2022;20:438.
33. Cobey KD, Haustein S, Brehaut J, et al. Community consensus on core open science practices to monitor in biomedicine. PLoS Biol. 2023;21:e3001949.
34. Waithira N, Mutinda B, Cheah PY. Data management and sharing policy: the first step towards promoting data sharing. BMC Med. 2019;17:80.