Journal of eScience Librarianship
putting the pieces together: theory and practice

Full-Length Paper

Applying Data Analytics and Visualization to Assessing the Research Impact of the Cancer Cell Biology (CCB) Program at the University of North Carolina at Chapel Hill

Fei Yu and Barrie E. Hayes

University of North Carolina at Chapel Hill, NC, USA

Abstract

Objective: The purpose of this paper is to report on a research impact assessment (RIA) project conducted by the Health Sciences Library (HSL) at the University of North Carolina at Chapel Hill (UNC-CH) for the Cancer Cell Biology (CCB) program in the institution's cancer center through bibliometric data analysis and visualization.

Methods: A total of 642 publications produced by the CCB researchers from 2010 to 2014 was used as the original dataset. After the citations of these publications were cleaned and standardized, the citations were imported into selected bibliometric and other tools for quantitative analysis and visualization.

Results: The CCB program at the UNC Lineberger Comprehensive Cancer Center had significant scientific output and citation impact in the examined five-year period, which was quantitatively measured not only by the total number of publications and citation counts, but also by comparative citation impact measures. In addition, the research collaboration network visualizations helped identify the most productive CCB researchers, the most highly cited CCB researchers, the research groups composed by co-authors, and the internal and external research partners. Further, the research topic visualizations confirmed the alignment of publication concentrations with the five core areas on which the CCB program has been focusing.

Conclusions: The bibliometric data analysis and visualizations produced for this project were able to provide quick insights to the administrators in terms of identified patterns, trends, and gaps of the supported research activities.

Introduction

Globally, there have been growing interest in assessing the research impact in the field of biomedical and health sciences, an area that is under constant pressure to justify the significant investments received (Penfield et al. 2014). Research impact assessment (RIA) helps academic institutes, research groups, and researchers with program planning, funding, hiring, and promotion decisions. A recent review (Milat, Bauman and Redman 2015) identified 16 conceptual frameworks or models used in health-related research domains to assess research impacts; among the different RIA approaches, bibliometric analysis was shown to be a core method embedded in most models. Bibliometrics is the quantitative analysis of the bibliographic information for publications, including measures such as total number of publications, citation counts, mean normalized citation score (MNCS), h-index, and measures of interdisciplinarity and specialization as well. Since any significant research activity usually results in publications, bibliometric analysis is a quick, direct, and economical way to measure research output and quality. In addition, researchers generally agree that bibliometric analysis generates objective, sophisticated, quantitative measures of research impact for scientific outputs (Agarwal et al. 2016; Rosas, Kagan and Schouten 2011). Even federal funding agencies such as the National Institutes of Health (NIH) have mandated publication tracking and reporting as a component for grants renewal (Schneider et al. 2017). Overall, bibliometric analysis helps both researchers and organizations demonstrate their research capacity, capabilities, and competency.

Bibliometric analysis can be conducted at multiple metric levels using available bibliometric products or tools. The two most authoritative citation database vendors, Elsevier and Clarivate Analytics (formerly known as Thomson Reuters), provide a series of bibliometric tools for multiple levels of measurement. For example, at the journal level, Elsevier offers SCImago Journal Rank (SJR), Source-Normalized Impact per Paper (SNIP), Impact Per Publication (IPP), and Citescore through its product, Scopus, while Clarivate Analytics offers Impact factor and Eigenfactor through its world-renowned product, Journal Citation Reports (JCR) (Plume and Colledge 2016; Clarivate Analytics 2017). At the author level, both vendors provide citation reports and h-index. Both vendors also offer citation counts and comparative citation impact metrics including Field-Weighted Citation Impact (FWCI) and Citation Benchmarking (CB) from Elsevier, and Category Normalized Citation Impact (CNCI) from Clarivate Analytics' InCites. In addition, Elsevier supplies PlumX altmetrics data through Scopus. At the institutional level, both vendors have products for impact benchmarking across institutions or research groups such as SciVal (Elsevier), Pure (Elsevier), and InCites (Clarivate Analytics). Beyond these proprietary bibliometric products, in 2016, the NIH Office of Portfolio Analysis launched a free web-based metrics application, iCite, to provide a new metric, Relative Citation Ratio (RCR), for citation benchmarking to NIH-funded research publications (NIH Office of Portfolio Analysis 2017). The RCR has quickly become a third comparative-citation impact indicator, joining FWCI (Elsevier) and CNCI (Clarivate Analytics).

Some of the bibliometric tools and products described above are primarily designed to track citation counts and analyze citation impact. However, citation analysis can be biased and controversial (Vucovich, Baker, and Jack 2008). For example, some works are cited because of errors and inaccuracies. In addition, citations accrue over time and older articles often have been cited more frequently than new ones. There are also citation rate variations across different disciplines. The application of citation analysis and bibliometrics more broadly in

evaluating the impact of a research unit's scientific output "requires a clear purpose, context, and full understanding of the limitations" (Rosas, Kagan and Schouten 2011, 2). In recent years, normalized impact scores for article-level measurement such as CNCI, FWCI, CB, and RCR have addressed some limitations in citation impact measurement and are regarded as a type of "better indicator of the influence of a set of papers than publication counts or average citation counts" (Schneider et al. 2017, 49). These normalized impact scores assess an article's relative influence by comparing with "similar" articles. For example, The FWCI and CB from Scopus take into account the following factors to ensure "apples are compared to apples:" the year of publication, document type, and the disciplines associated with its source (Colledge and Verlinde 2014). The RCRs and NIH percentile generated by iCite quantify the influence of an article by comparing it with the average NIH-funded publications in the same field and year (Hutchins et al. 2016).

Besides traditional research productivity and citation impact-based metrics, the scale of research collaboration has become a significant metric in evaluating the research outcomes in biomedical and health sciences. For instance, the community of clinical and translational awardees (CTSAs) have used social network analysis methods to assess co-authorship, institutional collaboration, and grant collaboration efforts with the aim of accelerating translational research (Sorensen, Seary and Riopelle 2010; Brian et al. 2013; Vacca, McCarty and Conlon 2015; Nagarajan et al. 2015). In these studies, locally-developed visualization tools were applied to collaboration analysis.

Commercial bibliometric products often incorporate some limited capabilities for bibliometric data visualization. Proprietary products like Scopus, Web of Science, InCites, and SciVal currently offer basic graphic display functions for search results analysis. At the article level, Scopus and Web of Science provide visualized statistics about publication year, source title, author, affiliation, country, document type, and subject area using line, bar, or pie charts (Beatty 2016; Web of Science 2017). Both the data and graphics can be easily exported. iCite automatically calculates the statistical indicators including the weighted RCR and the max, the mean, the median, and the standard error of the mean of the papers in the group for both citations per full calendar year after publication and the RCR (NIH Office of Portfolio Analysis 2017). Each publication's RCR and NIH percentile scores can also be directly downloaded as a detailed spreadsheet. At the institutional level, SciVal and InCites provide visualized citation impact benchmarking for comparing institutional research impact. However, the subscriptions to these bibliometric products are usually expensive.

The Health Sciences Library (HSL) at the University of North Carolina at Chapel Hill (UNC-CH) has long provided traditional bibliometric analysis services, upon request, to individual researchers and campus units. These bibliometric analysis services started to ramp up in 2012 as HSL librarians partnered with the institution's CTSA on their program evaluation efforts. In 2016, HSL began exploring citation network analysis and visualization tools to incorporate alongside bibliometric tools and methods in its work with the CTSA. Building on these explorations, HSL initiated an impact measurement and visualization (IMV) service as part of its "Research Hub" suite of researcher-targeted services. Since 2016, the requests for HSL IMV services from research units beyond the CTSA have grown significantly. One of the key research centers that HSL provided IMV services to was the UNC Lineberger Comprehensive Cancer Center. This paper discusses HSL's work with one of the cancer center programs as an example of the IMV services that HSL provides.

The Cancer Cell Biology (CCB) Program at the UNC Lineberger Comprehensive Cancer Center has been dedicated to the study of cancer-related basic and translational research in five core areas, including cell cycle regulation and tumor suppression, cell adhesion, cell signaling by growth factors and receptors, chromatin regulation and epigenetics, and angiogenesis and vascular biology (UNC Lineberger Comprehensive Cancer Center 2018). The administration of UNC-CH sought assistance from the HSL to develop an overview of research competence and achievement of its CCB program through bibliometric analysis.

The purpose of this project was to measure and demonstrate the research impact of the Cancer Cell Biology (CCB) publication output at UNC-CH through bibliometric data analysis and visualization. Both proprietary bibliometric products available at UNC-CH (i.e., Scopus) and free bibliometric tools (i.e., iCite & VOSviewer) were selected to investigate the following questions:

1. What is the quantitative scientific output and impact of the CCB program at UNC-CH?

2. What is the research collaboration status and scope of the CCB program at UNC-CH?

3. What research areas have CCB investigators focused/published on?

4. What quick insights can bibliometric analysis and visualization provide to CCB administrators?

Methods

Data source

Since citations need years to accumulate, to objectively assess the citation impact, a total of 642 publications produced by CCB researchers from 2010 to 2014 were used as the original dataset for this project. The cancer center at UNC-CH provided HSL with the brief citations of these 642 publications. For the convenience of describing the CCB program at UNC-CH and the dataset, they are referred to as "CCB program" and "CCB publications," respectively, in this article.

Tools

Scopus has much broader literature coverage in the biomedical and life sciences than its competitor, Web of Science (Falagas et al. 2008; Mongeon and Paul-Hus 2016). In addition, Scopus offers two comparative citation impact metrics, FWCI and CB with subscription. Despite not having a subscription to SciVal or InCites, UNC-CH's Scopus subscription allowed collection of the citation data needed for this project in June 2017. Further, iCite was adopted to provide RCRs. Last, Tableau was utilized for research productivity and citation impact comparison analysis, and VOSviewer (Version 1.6.5) (van Eck and Waltman 2010; van Eck and Waltman 2016), a free bibliometric network analysis tool was chosen for its collaboration network visualization and text mining functionality.

Workflow

First, HSL searched the Scopus database for the 642 CCB publications using the PMID, DOI, and Title fields. Bibliographic records were retrieved for 637 of these 642 publications. Then, the citations of the retrieved records were exported in full citation records format for citation

analysis. Second, an open source software – AutoIT (Version v3.3.14.2) (Bennett 2015) was used with a Perl script to scrape the values of FWCI and CB from Scopus for each of the 637 bibliographic records. Third, a total of 642 PMIDs were imported to NIH iCite. The generated NIH Relative Citation Ratios (RCRs) and NIH rank percentiles were downloaded in a spreadsheet for further analysis. Fourth, Excel was used to clean and standardize the downloaded Scopus citation data. Last, the cleaned and standardized data were imported to Tableau for quantitatively measurement of research productivity and citation impact comparison and to VOSviewer to generate collaboration networks and a research topic network.

Data Analysis & Visualization

To quantitatively assess the CCB scholarly output, this project focused on both article-level and institution-level metrics by examining traditional bibliometric measures (i.e., productivity, citation count, and average cites per paper) and relatively new measures (i.e., comparative citation impact, research collaboration, and research foci). Research productivity was measured by the number of publications in each calendar year. Tableau was used to visualize the publication productivity distribution. The citation count is an indicator of research influence. This project examined both the total and annual citation counts retrieved from Scopus for the CCB publications.

The average number of cites per paper is one of the important citation metrics and shows the average citation impact of each publication in a collection entity. This project used Elsevier's algorithm (Colledge and Verlinde 2014, 56) to compute the average cites per paper, which is to divide the total number of citation counts by the total number of publications for a defined period of time. Both the overall average cites per paper and the average cites per paper in each calendar year were calculated and visualized in Tableau.

The comparative citation impact demonstrates a publication's research influence by comparing it to "similar" articles. In this project, FWCI, CB, RCRs, and NIH Percentile of CCB publications were all examined for comparative citation impact. The comparison was visualized in Tableau.

The research collaboration impact was measured by a series of extracted collaboration networks such as co-author collaboration network, country collaboration network, and internal and external institutional collaboration networks. The co-author collaboration network was further analyzed by generating a co-authorship time overlay map and a citation impact density map. In addition, a research topic network map and a topic density map were created by analyzing the high occurrence key terms extracted from the title and abstract fields of CCB publications. All of the collaboration networks, the research topic network, and the topic density map were produced and visualized using VOSviewer.

Outcomes

Scholarly output – Productivity

With regards to research productivity, the CCB publications grew steadily from 2010 and reached a peak in 2013. Research productivity decreased slightly in 2014, the last publication year of analyzed publications. Over the five-year period, the average number of publications per year is 128, a substantial level of annual research output from the CCB program.

Figure 1: CCB Program Scholarly Publication Output by Year (Tableau).
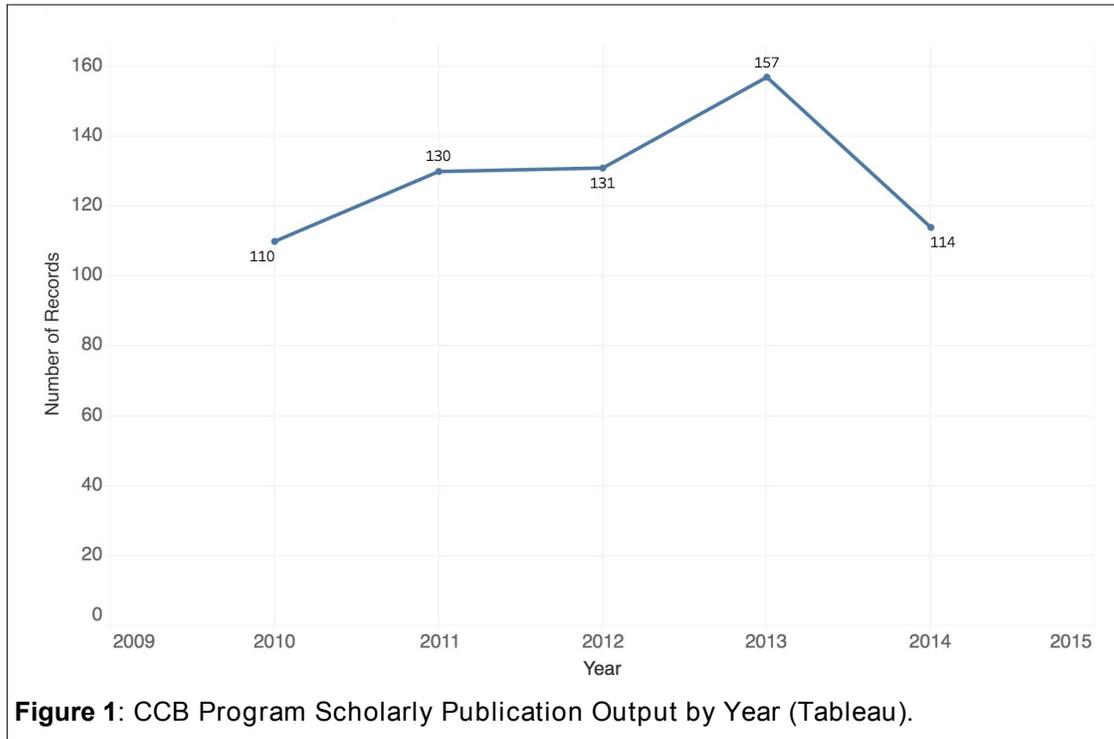
Scholarly output – Citation counts and impact

Scopus citation counts

Out of the total 642 CCB publications, 637 of them were retrieved from Scopus. The citation overview in Scopus shows 637 publications have accumulated 23,886 citation counts as of July 2017 (Figure 2). Particularly, one article (Ito et al. 2011) received 1,240 citation counts, which indicated significant research impact on peers. Overall, the citation counts of CCB publications have continued to increase over time since 2010.

Average cites per paper – Scopus vs. iCite

According to Scopus citation counts, on average, each of the 637 CCB publications has about 7.5 cites per paper over the examined five-year period. In iCite, 638 papers were matched. The iCite statistics shows the mean of "Cites/Year" of CCB publications is 5.81. The average cites per year for each CCB publication in Scopus is slightly higher than iCite. However, the annual distributions of average cites per paper from these two bibliometric sources followed similar patterns, which is a normal downward trend over time. This trend is expected because while the number of publications increased from 2010 to 2013, the citations need time to accrue, especially for more recent papers. Therefore, the annual average cites per paper showed the typical citation decline over time.

Comparative citation impact

In iCite, NIH-funded papers are the benchmark for RCR scores. For example, "a paper with an RCR of 1.0 has received the same number of cites per year as the average NIH-funded paper
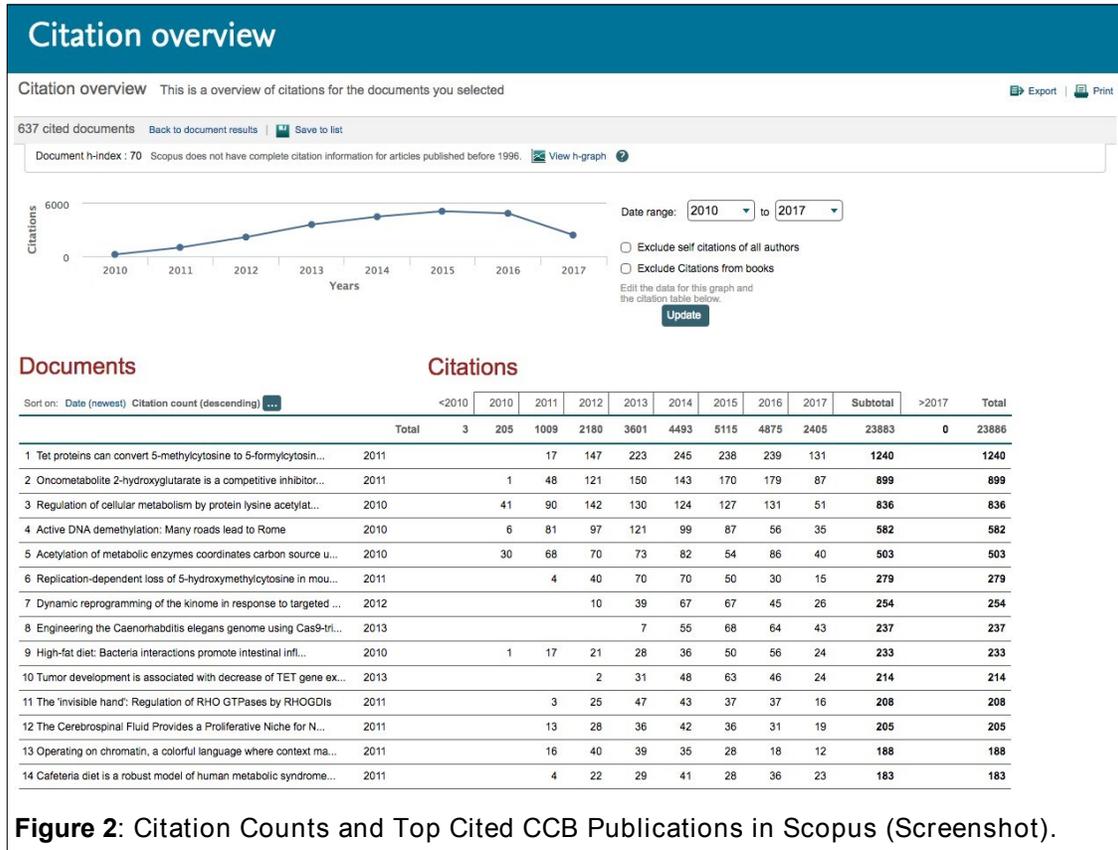
Citation overview

Citation overview This is a overview of citations for the documents you selected

Export | Print

637 cited documents | Back to document results | Save to list

Document h-index : 70 Scopus does not have complete citation information for articles published before 1996. View h-graph

Date range: 2010 to 2017

☐ Exclude self citations of all authors
☐ Exclude Citations from books

Edit the data for this graph and the citation table below.

Update

Documents / Citations

Sort on: Date (newest) Citation count (descending) ...

Documents	Year	<2010	2010	2011	2012	2013	2014	2015	2016	2017	Subtotal	>2017	Total
Total		3	205	1009	2180	3601	4493	5115	4875	2405	23883	0	23886
1 Tet proteins can convert 5-methylcytosine to 5-formylcytosin...	2011			17	147	223	245	238	239	131	1240		1240
2 Oncometabolite 2-hydroxyglutarate is a competitive inhibitor...	2011		1	48	121	150	143	170	179	87	899		899
3 Regulation of cellular metabolism by protein lysine acetylat...	2010		41	90	142	130	124	127	131	51	836		836
4 Active DNA demethylation: Many roads lead to Rome	2010		6	81	97	121	99	87	56	35	582		582
5 Acetylation of metabolic enzymes coordinates carbon source u...	2010		30	68	70	73	82	54	86	40	503		503
6 Replication-dependent loss of 5-hydroxymethylcytosine in mou...	2011			4	40	70	70	50	30	15	279		279
7 Dynamic reprogramming of the kinome in response to targeted ...	2012				10	39	67	67	45	26	254		254
8 Engineering the Caenorhabditis elegans genome using Cas9-tri...	2013					7	55	68	64	43	237		237
9 High-fat diet: Bacteria interactions promote intestinal infl...	2010		1	17	21	28	36	50	56	24	233		233
10 Tumor development is associated with decrease of TET gene ex...	2013				2	31	48	63	46	24	214		214
11 The 'invisible hand': Regulation of RHO GTPases by RHOGDIs	2011			3	25	47	43	37	37	16	208		208
12 The Cerebrospinal Fluid Provides a Proliferative Niche for N...	2011			13	28	36	42	36	31	19	205		205
13 Operating on chromatin, a colorful language where context ma...	2011			16	40	39	35	28	18	12	188		188
14 Cafeteria diet is a robust model of human metabolic syndrome...	2011			4	22	29	41	28	36	23	183		183

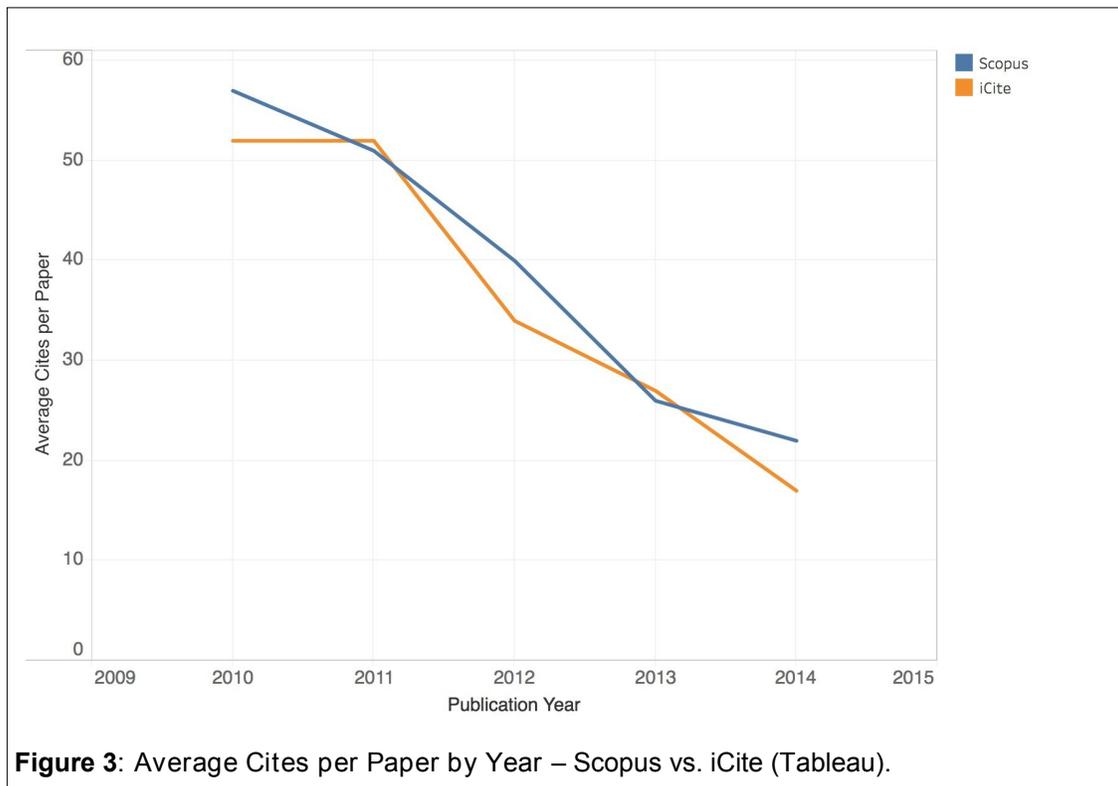**Figure 2**: Citation Counts and Top Cited CCB Publications in Scopus (Screenshot).

Figure 3: Average Cites per Paper by Year – Scopus vs. iCite (Tableau).

in its field, while a paper with an RCR of 2.0 has received twice as many cites per year as the average NIH-funded paper in its field" (NIH Office of Portfolio Analysis 2017). In this project, the mean of the RCRs of CCB publications was 1.8, which indicated CCB publications received almost twice as many cites per year as the average NIH-funded papers in the same fields during the time period 2010 to 2014. The annual distributions of two comparative citation impact metrics, RCR and FWCI, were compared in Figure 4. For both metrics, a score greater than 1.0 means CCB papers are cited more frequently than expected according to the average cites for similar papers. The patterns of the two metric score distributions approximately match each other. Consistent with previous studies, this project also found that Scopus database has higher citation counts than iCite which is likely due to its broader coverage of journal articles in biomedical and health sciences.

Figure 4: FWCI and RCR Comparison for CCB Publications (Tableau).
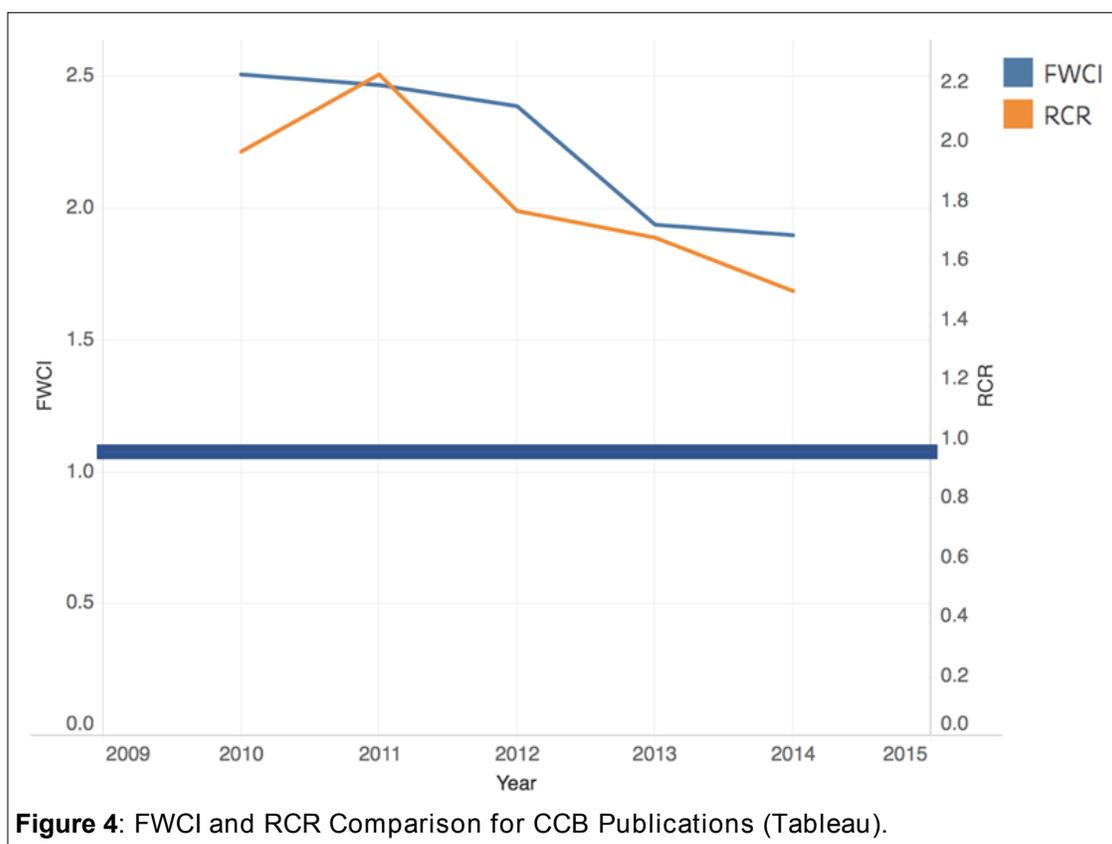
With respect to the Scopus Citation Benchmarking (CB), data analysis found that about 77% of CCB publications were more frequently cited compared to the average number of citations for articles published in the same time period, in the same discipline, and in the same document type (i.e., > 50th percentile). In addition, 16% of CCB publications were in the top 10% citation rank globally (i.e., >=90th percentile). In terms of the NIH percentile provided by iCite, 53% CCB publications were above NIH 50th percentile, which means slightly more than half of the CCB publications had superior citation impact compared to the average NIH-funded papers. Further, 9% CCB publications were in the top 10% NIH citation ranking (i.e., > = 90th percentile).

Figure 5: Scopus Citation Benchmarking and iCite NIH Percentile of CCB Publications (Tableau).
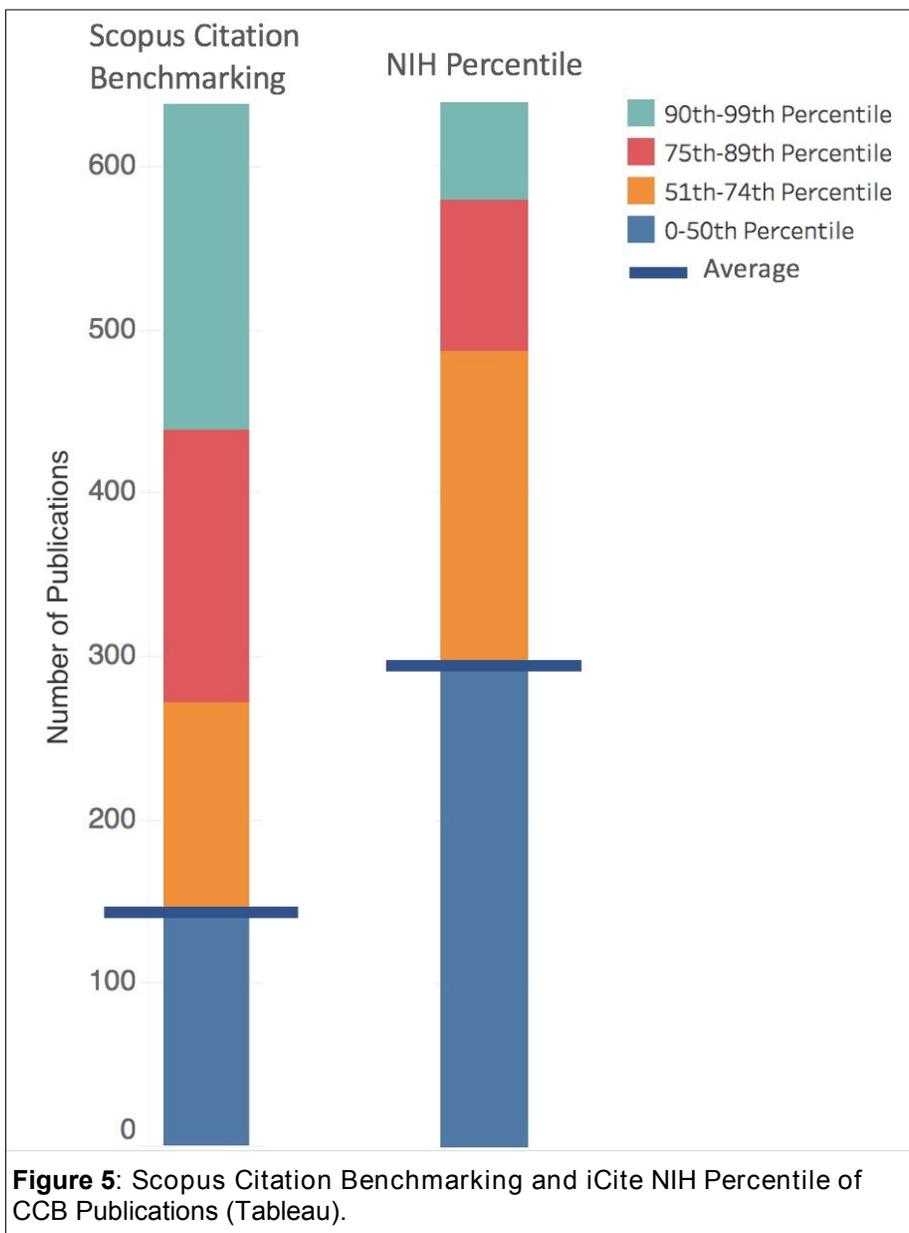
Co-author collaboration network visualization

Over 2,500 unique author names were extracted from CCB publications. Among them, 133 authors affiliated with UNC-CH contributed to at least five publications. In the co-authorship collaboration network, each author is represented by a node. The size of the node depends on the number of articles that an author has authored. For example, authors like Zhang Y., Xiong Y., and Clemmons D. are represented by large nodes, indicating their significant research productivity. This co-author collaboration network not only revealed the most productive CCB researchers in the analyzed publication collection, but also displayed the various research groups that these authors formed via differently colored clusters.
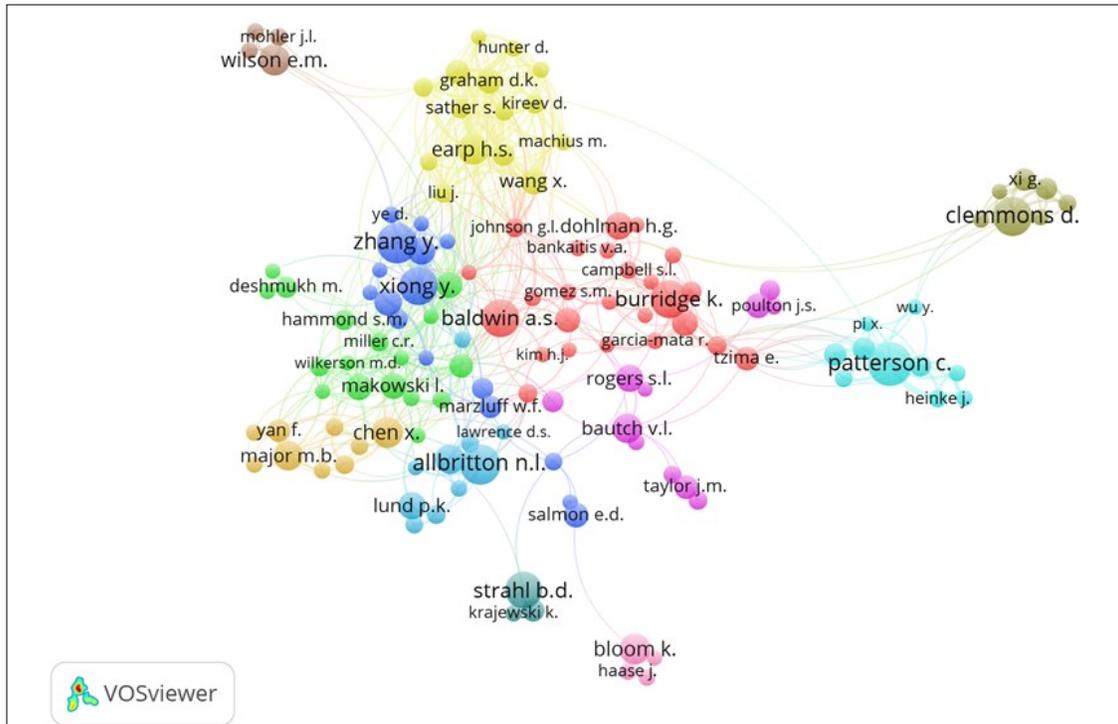
Figure 6: Co-author collaboration Network for CCB Authors Who Contributed to Five or More Articles, 2010-2014 (VOSviewer).
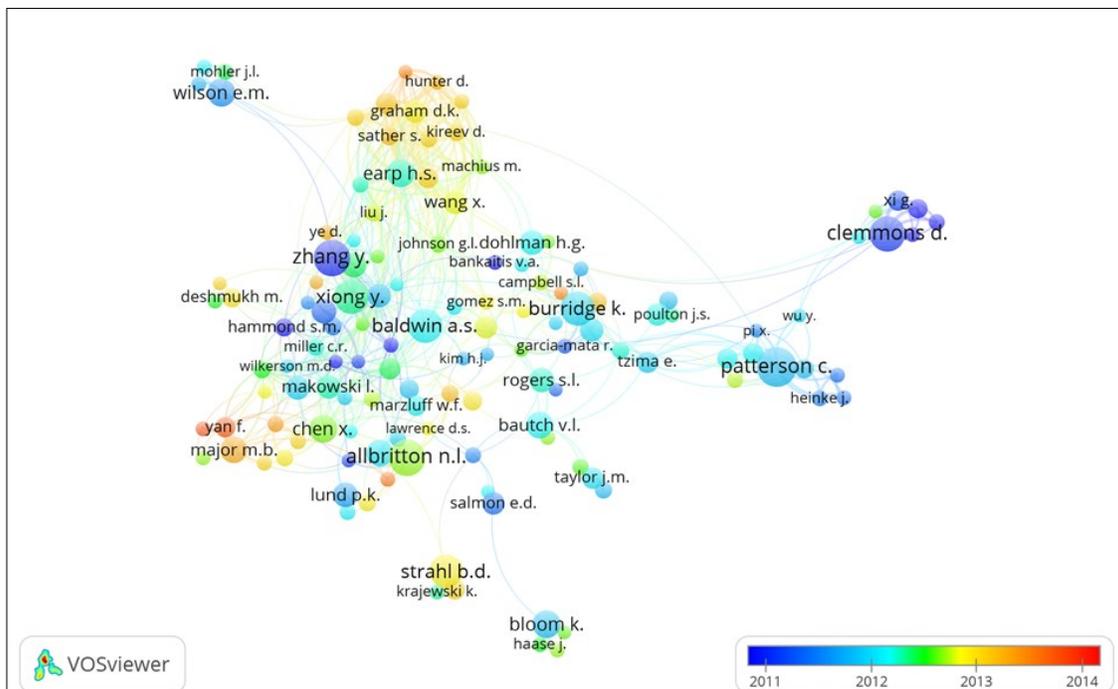

Figure 7: Co-author Collaboration Network with Time Overlay for CCB Publication Authors Contributing to Five or More Articles, 2010-2014 (VOSviewer).

The visualization of the co-authorship network with time overlay provides an overview of the key authors' research activities over time. For instance, the CCB publications produced by Clemmons, D., Patterson, C., or Zhang, Y. were mostly published before 2012 while the publications produced by Major, M.B., Yan, F., or Strahl, B.D. were published more recently.

This project also used the co-authorship citation density map to identify the most cited CCB authors. The authors in red were cited more frequently than those in yellow or green. The authors who were cited more frequently (e.g., Zhang, Y. and Xiong Y.) tend to have greater research influence on peers.

Figure 8: Co-author Citation Density Map for CCB Authors Contributing to Five or More Articles, 2010-2014 (VOSviewer).
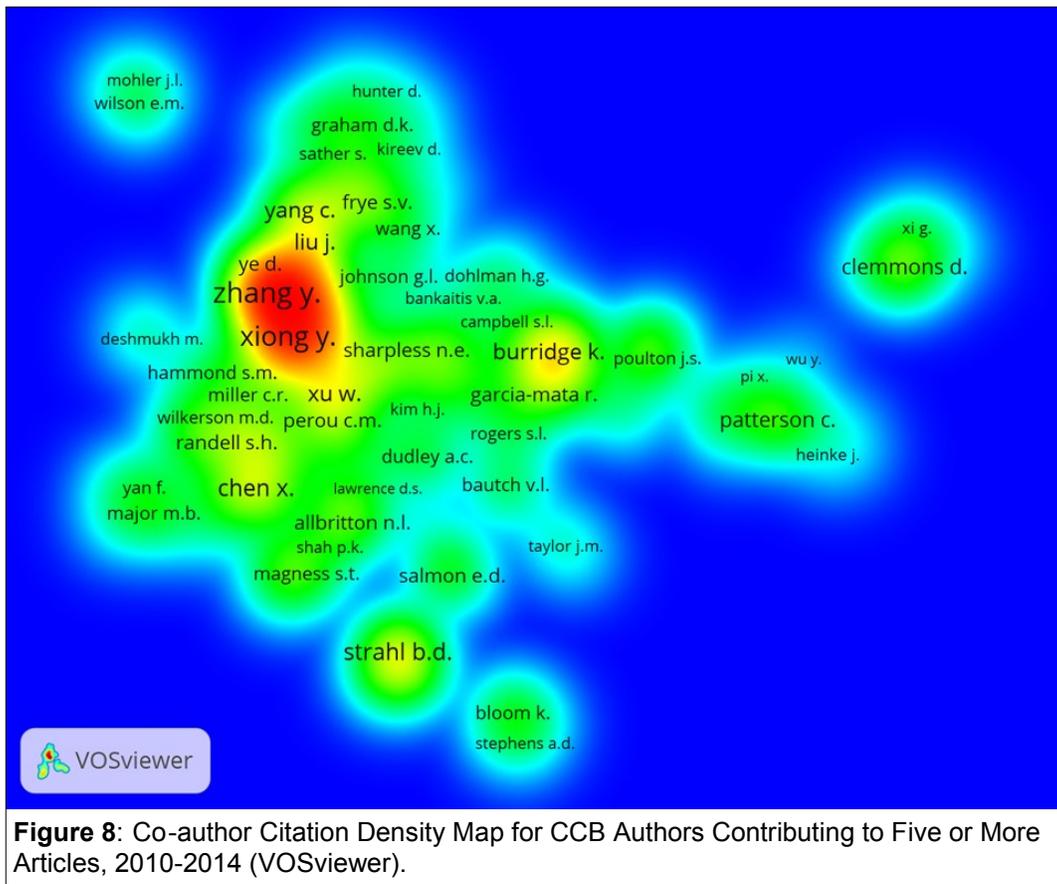
Country collaboration network visualization

CCB researchers collaborated with other researchers from 30 countries. The top collaborating countries include China, Germany, and Canada.

Internal organization collaboration network visualization

The internal organization collaboration analysis focused on the most granular level of UNC-CH units such as departments, centers, and divisions. More than 130 unique units were extracted from the CCB publications. The internal organization collaboration network showed a variety of internal units participated in the CCB studies. The top three are Department of Biology, Department of Biochemistry and Biophysics, and Department of Medicine.

Figure 9: Country Collaboration Network for CCB Authors (VOSviewer).

Figure 10: Internal Organization Collaboration Network for CCB Researchers, 2010-2014 (VOSviewer).

External organization collaboration network visualization

The external organization collaboration network analysis focused on the co-author organizations that were outside of UNC-CH at the macro level. For example, if an author's affiliation was stated at a granular level such as Harvard School of Medicine or Harvard School of Public Health, the affiliation data was then standardized to the macro organization level – Harvard University. Therefore, about 300 unique external organizations were identified through the author affiliation analysis. The external collaboration network was constructed based on the 100 most frequently collaborated external organizations. This network discovered that CCB researchers collaborated with peers from universities, research institutes, hospitals, clinics, industries, and government agencies at local, national, or international levels. The top external collaborating organizations included NC State University, Duke University, and East Carolina University. Internationally, the top collaborating organization was Fudan University in China. This external organization collaboration network demonstrated the broad reach of the CCB researchers.
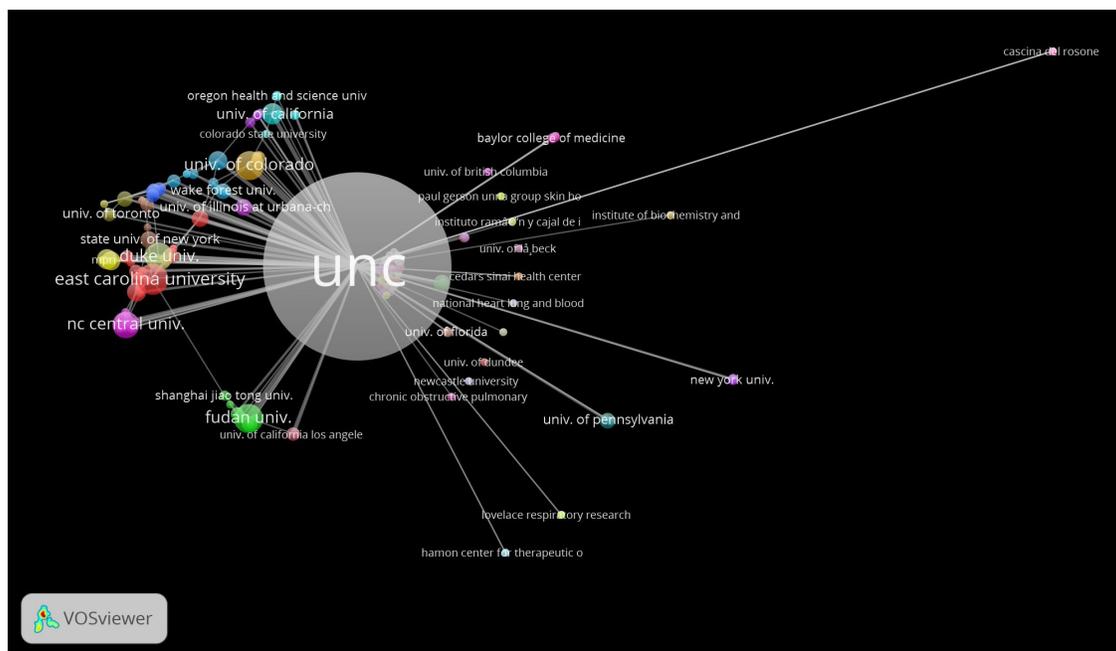
Figure 11: External Organization Collaboration Network for CCB Researchers, 2010-2014 (VOSviewer).

Topic mapping visualization

There were 280 high occurrence key terms (i.e., word terms occuring more than 10 times) extracted from both the title and the abstract field of CCB publications, and 160 of these key terms have strong relationships. Both the topic network map and the topic density map revealed that CCB studies emphasized the same five core areas that they stated publicly as their research foci such as cell cycle, cell adhesion, cell signaling by growth factors and receptors, the genesis, progression, and suppression of tumors, chromatin regulation and epigenetics, and angiogenesis.

Figure 12: Research Topic Network map for CCB Publications, 2010-2014 (VOSviewer).

Discussion

The bibliometric data analysis and visualization answered the questions that were outlined for investigation at the beginning of the project. First, the CCB program at UNC-CH produced significant scientific output and citation impact from 2010 to 2014, which was quantitatively measured not only by the total number of publications and citation counts, but also by comparative citation impact measures that are better indicators of the research influence of publications. Second, the co-author collaboration network visualizations (i.e., Figure 6, 7, & 8) helped identify the most productive CCB researchers, the most highly cited CCB researchers, and the network of co-authors that compose their research groups. The organization and country collaboration network visualizations (i.e., Figure 9, 10, & 11) revealed the landscape of CCB collaborations with internal and external partners, which CCB administrators found unexpectedly extensive. Third, both the research topic network visualization (Figure 12) and research topic density visualization (Figure 13) confirmed the alignment of publication concentrations with the five core areas on which the CCB program has been focusing. In addition, the visualization of key research topics provided a quick overview of the significant research activities conducted by CCB researchers during a selected five-year period and revealed the most heavily studied areas across the CCB program including cell signaling by growth factor, tumor progression, and cell cycle regulation. Fourth, the trends and patterns of the CCB research activities identified through visualizations provided ready insights into the program's scholarly productivity, citation impact, research collaboration and research focus

Figure 13: Research Topic Density Map for CCB Publications, 2010-2014 (VOSviewer).

areas. These insights can help the administrators develop strategies for creating more interdisciplinary collaboration opportunities, connecting researchers with appropriate peers and organizations, and better allocating resources and funds to important but less intensively explored cancer research areas.

In this project, HSL library staff leveraged both proprietary products (i.e., Scopus & Tableau) and free tools (i.e., iCite & VOSviewer) to conduct the bibliometric data analysis and visualization work. Specifically, Scopus was utilized to provide quick citation data analysis and graphic display, a capability which makes it an excellent tool to facilitate librarian bibliometric analysis services. As a free web application, iCite offers a customizable benchmarking feature to citation impact assessment. When analyzing NIH-supported studies, iCite is a preferred bibliometric tool because it, unlike others, compares the research impact with NIH-funded research publications. For quantitative analysis, Tableau, a powerful business analytics product, can also be very useful to bibliometric studies. In this project, Tableau was utilized to process hundreds of citation records and generate both statistics and visualizations in real time (i.e., Figure 1, 3, 4, & 5). As an effective and easy-to-use bibliometric network analysis tool, VOSviewer was designed to analyze and visualize co-citation, bibliographic coupling, co-authorship relations, and organization networks. Particularly, the text mining functionality it offers is key to topic mapping scientific literature. In this project, all the research collaboration

visualizations and research topic visualizations were generated with VOSviewer (i.e., Figure 6 -13).

This project took an exploratory bibliometric approach to RIA for a large biomedical and health science research institute by stressing both traditional measures (e.g., productivity, citation count) and relatively new measures (e.g., comparative citation impact, research collaboration). In addition, it is one of the few published projects at this writing which focuses on the article-level metrics by utilizing both proprietary bibliometric products and freely available tools. The freely available bibliometric tools and associated methods introduced in this project may be particularly helpful to resource-constrained libraries.

Nevertheless, there are several limitations in this project. First, the bibliometric approach to RIA usually depends on the access to bibliometric resources and tools. Reported outcomes can be generated using different bibliometric resources and would then be difficult to compare and benchmark across studies. For example, both Web of Science and Scopus provide citation counts and comparative citation impact indicators. However, both databases have different journal coverage and use different algorithms to categorize subject areas. Not only does the citation count vary for one article, but also the top percentile publications in one resource may not be the top percentile publications in the other. Second, subscriptions to proprietary bibliometric products are expensive. Many institutions do not have access to all of the authoritative resources and products due to budget constraints. Therefore, some citation analysis activities cannot be performed. For instance, this project could only perform RIA analysis within one institution because UNC-CH does not subscribe to either SciVal or InCites which provide aggregated institutional-level citation benchmarking. Third, this project only conducted bibliometric analysis at the article level. Other measures such as the journal-level metrics were not addressed. Fourth, as discussed earlier, citation analysis can be biased and only offers approximate information about the scientific impact of publications, authors, or institutions (Waltman and Noyons 2017). Administrators should not make research management decisions based solely on bibliometric analysis. Citation impact is best combined with other metrics to provide a multi-faceted view of research impact, including measures like altmetrics, peer reviews and intellectual property products produced.

Conclusion

This project applied bibliometric data analytics and visualization to assessing the research impact by using selected proprietary and free tools. Specifically, the visualized comparative citation impact and the collaboration networks highlighted the high-quality impact and broad research reach of the CCB program at UNC-CH. Overall, the bibliometric visualizations greatly facilitated and enhanced the understanding of the research impact assessment and provided quick insights to the administration for decision making in terms of research management.

To build capacity to support the growing service demands and take on larger institutional projects, HSL expanded the IMV team in fall 2017 by drawing library staff and graduate research assistants from both the public services and the health technology and informatics units. With this cross-departmental HSL IMV team in place, the IMV service is now being more actively promoted on the library website and in HSL director and staff communications with health sciences department heads, research center directors, research office administrators, and other campus constituencies. As with the CCB project reported here, the team continues

to learn and leverage a combination of its subscription bibliometric resources and freely available tools to produce bibliographic analyses and visualizations that provide insights into the requesting programs' research activities and impact at the University of North Carolina at Chapel Hill.

Acknowledgments

Disclosure

The authors report no conflict of interest.

References

Agarwal, Ashok, Danmayanthi Durairajanayagam, Sindhuja Tatagari, Sandro Esteves, Avi Harley, Ralf Henkel, Shubhadeep Roychoudhury et al. 2016. "Bibliometrics: tracking research impact by selecting the appropriate metrics." *Asian Journal Andrology* 18(2): 296-309. http://www.dx.doi.org/10.4103/1008-682X.171582

Bennett, Jonathan. 2015. AutoIt (Version 3.3.14.2) [Software]. Available from https://www.autoitscript.com/site

Beatty, Susannah. 2015. "New Scopus Article Metrics: A better way to benchmark articles." *Elsevier Scopus Blog*. Last modified July 27. https://blog.scopus.com/posts/new-scopus-article-metrics-a-better-way-to-benchmark-articles

Beatty, Susannah. 2016. "Analysis tools." *Elsevier Scopus Blog.* Last modified May 23. https://blog.scopus.com/topics/analysis-tools

Brian, Jiang, Mengjun Xie, Umit Topaloglu, Teresa Hudson, and William Hogan. 2013. "Understanding biomedical research collaborations through social network analysis: A case study." In *2013 IEEE International Conference on Bioinformatics and Biomedicine*, Shanghai, China. 9-16. http://www.dx.doi.org/10.1109/BIBM.2013.6732728

Brian, Jiang, Mengjun Xie, Umit Topaloglu, Teresa Hudson, Hari Eswaran and William Hogan. 2014. "Social network analysis of biomedical research collaboration networks in a CTSA institution." *Journal of Biomedical Informatics* 52: 130-140. https://doi.org/10.1016/j.jbi.2014.01.015

Colledge, Lisa and Reinder Verlinde. 2014. "SciVal Metric Guidebook." *Elsevier*. https://www.elsevier.com/research-intelligence/resource-library/scival-metrics-guidebook

Clarivate Analytics. 2017. "World's most influential journals for 2017 unveiled in the journal Citation Reports from Clarivate Analytics." Accessed May 27. https://clarivate.com/blog/news/worlds-influential-journals-2017-unveiled-journal-citation-reports-clarivate-analytics

Falagas, Matthew, Eleni Pitsouni, George Malietzis and Georgios Pappas. 2008. "Comparison of PubMed, Scopus, Web of Science, and Google Scholar: Strengths and Weakness." *The FASEB Journal* 22(2): 338-342. http://dx.doi.org/10.1096/fj.07-9492LSF

Harris, Paul A., Jacqueline Kirby, Jonathan Swafford, Terri Edwards, Minhua Zhang, Tonya R. Yarbrough, Lynda Lane, Tara Helmer, Gordon Bernard and Jill Pulley. 2016. "Tackling the "So What" Problem in Scientific Research: A Systems-Based Approach to Resource and Publication Tracking." *Academic Medicine* 90(8): 1043-1050. http://dx.doi.org/10.1097/ACM.0000000000000732

Hutchins, B. Ian, Xin Yuan, James M. Anderson and George M. Santangelo. 2016. "Relative Citation Ratio (RCR): A New Metric That Uses Citation Rates to Measure Influence at the Article Level." *PLoS Biology* 14(9): e1002541. http://dx.doi.org/10.1371/journal.pbio.1002541

Ito, Shinsuke, Li Shen, Qing Dai, Susan C. Wu, Leonard B. Collins, James A. Swenberg, Chuan He, and Yi Zhang. 2011. "Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine." *Science* 333 (6047): 1300-1303. http://dx.doi.org/10.1126/science.1210597

Milat, Andrew J., Adrian E. Bauman and Sally Redman. 2015. "A narrative review of research impact assessment models and methods." *Health Research Policy and Systems* 13(18): 213-228. http://dx.doi.org/10.1186/s12961-015-0003-1

Mongeon, Philippe and Adèle Paul-hus. 2016. "The Journal Coverage of Scopus and Web of Science: A Comparative Analysis." *Scientometrics* 106(1): 213-228. http://dx.doi.org/10.1007/s11192-015-1765-5

Nagarajan, Radhakrishnan, Charlotte A. Peterson, Jane S. Lowe, Stephen W. Wyatt, Timothy S. Tracy and Philip A. Kern. 2015. "Social Network Analysis to Assess the Impact of the CTSA on Biomedical Research Grant Collaboration." *Clinical and Translational Science* 8(2): 150-154. http://dx.doi.org/10.1111/cts.12247

NIH Office of Portfolio Analysis. 2017. "iCite, New Analysis" https://icite.od.nih.gov/analysis

Penfield, Teresa, Matthew J. Baker, Rosa Scoble and Michael C. Wykes. 2014. "Assessment, Evaluations, and Definitions of Research Impact: A Review." *Research Evaluation* 23(1): 21-32. http://dx.doi.org/10.1093/reseval/rvt021

Plume, Andrew and Lisa Colledge. 2016. "New metrics will make journal assessment more complete and transparent." *Elsevier Connect*. https://www.elsevier.com/connect/new-metrics-will-make-journal-evaluation-easier-and-more-transparent

Rosas, Scott R., Jonathan M. Kagan, Jeffrey T. Schouten, Perry A. Slack and William M. K. Trochim. 2011. "Evaluating Research and Impact: A Bibliometric Analysis of Research by the NIH/NIAID HIV/AIDS Clinical Trials Networks." *PLoS ONE* 6(3): e17428. http://dx.doi.org/10.1371/journal.pone.0017428

Schneider, M., C. M. Kane, J. Rainwater, L. Guerrero, G. Tong, S. R. Desai and W. Trochim. 2017. "Feasibility of Common Bibliometrics in Evaluating Translational Science." *Journal of Clinical and Translational Science* 1(1): 45-52. http://dx.doi.org/10.1017/cts.2016.8

Sorensen, Aaron, Andrew Seary and Kenneth Riopelle. 2010. "Alzheimer's Disease Research: A COIN Study Using Co-authorship Network Analytics." *Procedia - Social and Behavioral Sciences* 2(4): 6582-6586. https://doi.org/10.1016/j.sbspro.2010.04.068

UNC Lineberger Comprehensive Cancer Center. 2018. "Cancer Cell Biology." Accessed January 16. https://unclineberger.org/research/research-programs/cancer-cell-biology

Vacca, Raffaele, Christopher Mccarty, Michael Conlon and David R. Nelson. 2015. "Designing a CTSA-Based Social Network Intervention to Foster Cross-Disciplinary Team Science." *Clinical and Translational Science* 8(4): 281-289. http://dx.doi.org/10.1111/cts.12267

van Eck, Nees Jan and Ludo Waltman. 2010. "Software survey: VOSviewer, a computer program for bibliometric mapping." *Scientometrics* 84(2): 523-538. https://doi.org/10.1007/s11192-009-0146-3

van Eck, Nees Jan and Ludo Waltman. 2016. VOSviewer (version 1.6.5) [Software]. Available from http://www.vosviewer.com

Vucovich, Lee A., Jason Blaine Baker and Jack T. Smith, Jr. 2008. "Analyzing the Impact of an Author's Publications." *Journal of Medical Library Association* 96(1): 63-66. http://dx.doi.org/10.3163/1536-5050.96.1.63

Waltman, Ludo, and Ed Noyons. 2017. *Bibliometrics for Research Management and Research Evaluation*. Leiden: CWTS BV.

Web of Science. 2017. "Web of Science Research Tools." Last modified November 30. https://images.webofknowledge.com/images/help/WOS/hs_research_tools.html