A new approach to evaluate scientific production

Short title

TeamTree analysis

Frank W. Pfrieger*

Centre National de la Recherche Scientifique, Université de Strasbourg, Institut des Neurosciences Cellulaires et Intégratives, Strasbourg, France.

https://orcid.org/0000-0001-7085-1431

*Corresponding author

Email: fw-pfrieger@gmx.de or frank.pfrieger@unistra.fr (FWP)

Abstract

Advances in science and technology depend on the work of research teams and the dissemination of their results through peer-reviewed articles, which constitute a fastly growing socio-economic resource. Current methods to mine publications in a field of interest focus on content, but the workforce credited by authorship remains largely unexplored. This hurdle can be overcome by a new bibliometric approach that visualizes the publication records of authors working in a user-defined field, delineates their genealogic and collaborative connections and provides citation-independent performance measures. This team-centered approach complements existing tools to mine the scientific literature across disciplines.

Keywords

Authorship, citation, collaborations, data science, genealogy, meta research, science of science, scientometrics, informetrics, bibliometrics, team science, TeamTree

Introduction

Progress in science and technology (S&T) depends on research teams working in specific fields of interest and on the publication of their results in peer-reviewed articles. Authorship on these articles credits contributions of individual team members with diverse expertise and skills [1, 2]. The rapidly growing body of scientific information [3] reflects past and current states of the art and represents an invaluable socio-economic resource guiding future research activities, policies and investments [4-9]. Its utility relies on refined methods to search and analyse scientific articles. Nowadays, computational algorithms processing large datasets inform about natural processes, help to formulate new hypotheses or to find new drugs, and

provide new insight in the roles of authors, teams and collaborator networks [10-26]. A typical user interested in a specific topic (e.g. physical phenomenon, chemical compound or human disease) queries a bibliographic database and scrutinizes the list of publications. However, the teams working in the user-defined field remain largely inaccessible. The present work aimed to overcome this hurdle. To this end, a new bibliometric approach and visuals were developed that give users interested in a specific field an instant view on the workforce based on a list of publications. The approach, further referred to as TeamTree analysis (TTA), is introduced here using selected research fields from different scientific disciplines as examples.

Materials and methods

TTA was implemented with the open source software R [27] and selected R packages (data.table [28]; dunn.test [29]; eulerr [30]; ggfortify [31]; ggplot2 [32]; ggrepel [33]; igraph [34]; plot3D [35]). The R script is available upon request to the author and at https://github.com/fw-pfrieger/TeamTree. Publications were obtained by queries of bibliographic databases (Table 1).

Table 1. Selected research fields and authors subjected to TTA.

Field / Discipline	Database	Pubs / Authors / Year
Aplysia	PubMed	4738 / 1624 / 1898
Circadian Clock[1] / Biomedicine	PubMed	17514 / 6759 / 1960
Chirped Laser Pulses[2] / Physics TS=chirped laser pulses	WOS	9021 / 4251 / 1968
Cosmic Inflation[3] / Astronomy	WOS	3311 / 1790 / 1981

TS=("cosmic inflation" OR "inflationary universe")		
CRISPR (clustered regularly interspaced short palindromic repeats OR crispr*)	PubMed	23689 / 12013 / 2002
Ice Core Climate[4] / Geoscience TS=ice core climate	WOS	9723 / 5904 / 1956
Organoids	PubMed	14610 / 10077 / 1946
Quantum Computer[5] / Computer Science TS=("quantum computer" OR "quantum computing")	WOS	28597 / 10390 / 1985
Supramolecular Chemistry[6] / Chemistry TS=supramolecular chemistry	WOS	33642 / 13061 / 1967
Gregory Y. Lip / Selected Team / Biomedicine	PubMed	3017 / 587 / 1991
Hagop Kantarjian / Selected Team / Biomedicine	PubMed	2195 / 366 / 1981
Didier Raoult / Selected Team / Biomedicine	PubMed	3034 / 342 / 1979

Summary of selected fields and authors, their discipline, the query term, the relevant bibliographic database, the number of publications and of authors and the first year of publication. Selected awards: [1]Nobel Prize Physiology/Medicine (2017); [2]Nobel Prize Physics (2018); [3]Kavli Prize (2014); [4]Vetlesen Award (2013); [5]Micius Quantum Prize (2018, 2019); [6]Nobel Prize Chemistry (1987).

TTA extracts a non-redundant list of last authors from articles with at least one additional author. Author names include initials to reduce author ambiguity. Single author articles count towards the publication record of last authors. For simplicity, last authors are further referred to as "authors". TTA assigns a chronologic index with alternating sign (AI; Table 2) to each author according to the year of the first publication as last author. Further, TTA attributes

randomly a color to each author and calculates several parameters summarizing each author's publication record, genealogy and collaborations (Table 2). Genealogic relations are based on offspring - ancestor pairs, where offspring is defined as authors, who appeared initially as first authors on a publication of an ancestor author. Accordingly, TTA calculates the author generation (AG; Table 2). Each offspring (AG = i+1) is assigned to the ancestor (AG = i) with the earliest common publication. A family is defined as progeny of a first ancestor (AG = 1) encompassing all offspring (AG > 1). TTA derives collaborations based on co-authorship [36] separating out- and in-degree connection, where the author lists other authors as co-authors and where the author is listed as co-author, respectively. Statistical analyses were performed using indicated tests. For comparison, citation-based measures of randomly selected authors were obtained manually from the Web of Science database selecting "articles" and excluding self-citations.

Table 2. Summary of TTA-derived parameters.

Parameter	Description
AC	Number of authors per publication
AG	Generation of an author, where AG ancestor = i and AG offspring = i+1
AI	Chronologic index of an author based on the first (last author) publication with alternating sign from one author ($+AI_n$) to the next ($-AI_{n+1}$) enabling a tree-like display of the workforce
CC	Collaboration count of an author: "out", number of co-authors; "in" number of authors that listed the author as co-author
FS	Family size: number of all progeny of a first generation ancestor
OC	Offspring count of an author: number of first authors on an author's publications that subsequently publish as last author
PCannu	Mean annual count of last author publications

PCcol	Number of publications with collaborators: "out", number of articles where the author is last author and a collaborator is listed as co-author; "in", number of papers where the author listed as co-author. Only articles with three authors or more are taken into account.
PCfirst	Number of publications as first author
PClast	Number of publications as last author including single-author articles
PCoff	Number of last author publications with offspring
TTP	TeamTree product: result of PClast × OC × CC (in+out)

Results

The features of TTA are introduced using selected fields of research from distinct disciplines of S&T (Table 1).

Visualizing the workforce driving research fields

TTA introduces a new type of visual, further referred to as TeamTree graph, to display the workforce driving a user-defined research field of interest based on the author lists of scientific articles and their years of publication. To this end, the publication years of each author are plotted against AI creating a tree-like image. To exemplify this feature, TTA was applied to publications related to three fields of biomedical research that were chosen because of their distinct dynamics (Fig 1; Table 1). The corresponding TeamTree graphs reveal the long history and stable workforce of the *Aplysia* field [37], the fastly growing number of teams working on CRISPR-related topics [38] and the peculiar development of the organoid field [39] showing distinct phases of workforce expansion (Fig 1).

Fig 1. TeamTree graphs showing the development of the workforce in selected fields of biomedicine.

TeamTree graphs reveal the distinct duration, size and growth of the three fields shown as examples (Table 1). They are generated by plotting the years of publication of each author against AI. Grey lines connect publication years of each author. Circle area represents PClast per year.

Display and quantitative analysis of publication record, genealogy and collaborations

To further illustrate its features, TTA was applied to publications related to "circadian clock" (Clock), a representative field of biomedical research [40] (Table 1). Figure 2 shows results with respect to the publication records of authors in the field. Individual authors published as many as 115 articles (PClast), but two thirds of the workforce (69%) contributed single articles (Fig 2B) even after exclusion of authors entering during the last two years (67%). The Clock field expanded exponentially within the last decades as indicated by linearly growing annual counts of newly entering authors and of published articles per year, respectively (Fig 2C). Ranking authors by PClast values identified the top contributors to the field with respect to publication record (Fig 2D).

Fig 2. Publication record in the Clock field.

(A) TeamTree graph representing the publication records of authors working in the Clock field. Circle area, PClast per year. (B) Left, publication counts per author separating last and first author articles. Circle area, PCannu. Right, relative frequency distributions of PC values shown on the left. (C) Number of authors entering the field per year (orange) and of articles (black) published per year. (D) Authors with top ten PClast values indicated by circle area.

TTA revealed ancestor–offspring relations in the Clock field based on last author–first author pairs of PubMed articles, respectively (Fig 3). A quarter of authors published previously as first authors thus qualifying as offspring in this field (Fig 2B), whereas 13% of the authors qualified as ancestors. They generated up to 24 offspring and published up to 71 articles with their offspring (Fig 3B). Overall, the Clock field comprised 591 families with a variable number of members spanning up to 8 generations (Fig 3B). For the last two decades offspring authors and publications with offspring represented a small, but constant fraction of the workforce entering the field each year and of the annual scientific production (Fig 3C). Ranking authors by OC values indicated the most prolific authors and their families (Fig 3D).

Fig 3. Genealogic relations in the Clock field.

(A) TeamTree graph showing ancestor-to-offspring connections among authors. Connections of authors with top ten OC values are shown in color (names indicated in panel D). Circle area, OC. AI signs of offspring and of ancestors were adjusted. (B) Left, from top to bottom, OC, PCoff, AG and FS values. Circle area, PCannu. Right, Relative frequency distributions of parameters shown on the left. (C) Fraction of new offspring authors (orange) and of publications with offspring (black) compared to total number per year. (D) Names and family connections of authors with top ten OC values indicated by circle area.

Collaborations in the Clock field were delineated based on co-authorship on PubMed articles (Table 2; Fig 4). In total, 53% of the authors in the Clock field established a variable number of out- and in-degree collaborations and published collaborative papers as last and co-author, respectively (Fig 4B). During the last two decades, collaborators represented half of the new authors entering per year and their contribution remained fairly constant (Fig 4C), whereas

the number of authors per article increased steadily (Fig 4A). Ranking authors based on connection counts revealed the most strongly connected players in the field. They formed a large network with a diameter of 8, a mean distance of 3.82 and an edge density of 0.0117 (Fig 4D).

Fig 4. Collaborations in the Clock field.

(A) TeamTree graph showing collaborations between last authors (out; negative AI) and co-authors (in; positive AI). Connections of authors with top ten CC values (in+out) are shown in color. Circle area, CCout and CCin values of these authors. Inset, mean AC per article published each year. (B) Left, counts of respective collaborators and of resulting articles per author. Circle area, PCannu. Right, relative frequency distributions of parameters shown on the left. (C) Fractions of new collaborating authors (orange) and of collaborative publications (black) compared to total numbers per year. (D) Names of authors with top ten CC values (in+out) and their networks. Circles and squares, offspring and non-offspring authors, respectively. Circle area, CC (in+out) values normalized to the maximum.

Workforce dynamics and field development

TTA was used to explore how the workforce of the Clock field developed over time. Plotting the number of authors entering and exiting the field based on the first and last year of their publications, respectively, revealed a strong growth of the workforce up until three years ago (Fig 5). The publication periods or life-span of authors ranged up to five decades, but the large majority published only during one year and in most cases a single article (Fig 2C; Fig 5A-C). Separating "Newcomers" entering the field per year from "Established" authors revealed that the established workforce consisted mostly of authors with genealogic and

collaborative connections, whereas most newcomers had collaborative connections or no ties at all and contributed single articles (Fig 5D).

Fig 5. Workforce dynamics in the Clock field.

(A) Annual counts of authors entering (green bars) and leaving the field (red bars). Lines show cumulative sums. (B) Publication periods of individual authors in years. (C) Bars and lines showing the relative frequencies of all publication periods and the cumulative relative frequencies of publication periods from indicated categories, respectively. Col, authors with collaborative but no genealogic connections; Off, genealogic but no collaborative connections; Off+Col, both types of connections; Rest, without connections. Statistically significant differences among groups are indicated (Kruskal-Wallis tests chi-squared = 267.1734, df = 3, p = 0. Asterisks: *, $p < 0.05$; **, $p < 0.01$; ***, p = 0; post-hoc Dunn test, Benjamini-Hochberg adjusted; group size = 301; adjusted to minimum by random selection). (D) Annual author (filled) and publication counts (white) of newcomers (left) and established teams (right) from the indicated categories. In each category, authors with single publications are represented by grey bars. Scale bar indicates number of authors or publications.

Author contribution based on publication output, offspring and collaborations

A recurrent goal of bibliometric analyses is to gauge author contributions to a field of research, as this can help to choose referees/experts, recruits and collaborators. TTA delivers three parameters (PC, OC and CC) to estimate the contribution of authors as illustrated in Figs 2-4. Intersection of the top 100 authors for each parameter revealed overlap between PClast-, OC- and CCin+out-based rankings. A core of 35 authors figured among the top in all three categories (Fig 6). Three-dimensional plots of these parameters revealed that authors

occupy distinct volumes (Fig 6B) indicating that the product of the three measures, further referred to as TeamTree product (TTP; Table 2), enables a more differentiated author ranking than each parameter alone (Fig 6C). Notably, TTP values of individual authors correlated significantly with numbers of citing articles (rho = 0.88; $p = 2.2\ 10^{-16}$) and with H indices (rho = 0.94; $p = 2.2\ 10^{-16}$; Spearman's rank correlation; Fig 6D). These findings indicated that TTP can serve as a citation-independent measure of an author's contribution to a field. Teams with highest TTP values included Nobel prize winners (M. Rosbash: 7th, M.W. Young: 18th, J.C. Hall: 41st out of 6759; Table 1; Fig 6C; S1 File).

Fig 6. Introduction of TTP as a new measure of scientific production.

(A) Numbers of intersecting authors ranking among top 100 for each parameter (PClast, OC, CCin+out). (B) Scatterplot of indicated parameters for authors with top ten TTP values. (C) Top, TeamTree graph showing TTP values of authors (circle size) normalized to maximum. Colored circles and names, authors with top ten TTP values. Bottom, log10(TTP) values of authors. (D) Number of citing articles (top) and H index (bottom) of individual authors plotted against their TTP values.

Analysis of distinct scientific disciplines and of authors with exceptional publication records

TTA reveals the workforce driving a field of interest. The choice is not limited to biomedicine. To illustrate this point, TTA was applied to representative fields in other disciplines (Table 1). Combined TeamTree graphs summarizing the publication record, genealogic and collaborative connections and TTP values are shown in Fig 7. In addition, TTA allows to track the activity of single authors, for example those with exceptional publication counts [41], and to separate articles originating from offspring and collaboration.

Thus, TTA can providing a more detailed view on a specific author's publication record (Table 1; Fig 7B-D).

Fig 7. TTA of selected fields in different disciplines and of selected authors with exceptional publication records.

(A) Combined TeamTree graphs summarizing the indicated parameters (PClast, OC, CCin+out, TTP) for selected research fields. Only authors with top ten values for each parameter are shown. Circle areas, relative counts normalized to respective maximum. Names, authors with top ten TTP values. Horizontal and vertical scale bars, number of authors and years, respectively. Stars indicate awardees (Table 1). (B) TeamTree graphs showing publication counts of indicated authors (Table 1). (C) Publication rates per year of the same authors separating articles by contributions (First, first author; Col in, in-degree collaborations; Col out±Off, out-degree collaborations with and without offspring; Off–Col out, offspring without out-degree collaborators; Rest, all other publications). Insets, contributions of indicated categories to total publication record. (D) Mean AC per article published each year.

Discussion

TTA gives users an instant view on the workforce driving a field of interest based on authors and publication years of scientific articles. The new TeamTree graphs depict the origin, size and development of a research field and show the genealogic and collaborative connections among authors. TTA can be performed using a freely available R script. An open source project aims to provide access to a web-based TTA with enhanced functions and visuals.

TTA reveals how the workforce impacts the development of research areas. Only a fraction of authors contributed for more than one year to the Clock field. This supports

previous findings that in many research areas only a small fraction of the workforce publishes during long periods of time [42]. Genealogic and collaborative connections prolonged the life-span of teams within the Clock field. These observations are in line with previous studies highlighting the relevance of training and mentorship [43-45] and the importance of collaborations [5, 46-49]. TTP represents a new measure to estimate the contribution of authors to a specific field. This parameter takes into account the publication record, training/mentorship and cooperation. Stratified publication counts separating last author, first author and co-author contributions provide a detailed view on the publication record. Offspring counts value training and mentoring of junior scientists that continue their career within the field. Collaboration counts signify recognition within the field due to specific expertise and capacities. TTP is citation-independent, but scales with the citation-based H index [50] and thereby complements other bibliometric measures [6, 51-56]. A few caveats should be mentioned. TTP values are field-specific, they scale with the size of research groups and they depend on the life-span of authors in the research field. Therefore, the TTP-based ranking is context-dependent and unsuited for junior scientists [57] and the metric cannot gauge innovative, ground-breaking contributions from small teams or from teams that contribute only briefly to a field. Honorary and ghost authorship will confound results depending on their prevalence in the field of interest [58, 59]. Other limitations may be addressed in future versions. The automatic delineation of family connections from first author-last author pairs provides an alternative to efforts requiring user input [60-62]) (see also Mathematics Genealogy Project: https://www.genealogy.math.ndsu.nodak.edu/; Academic family tree: https://academictree.org/). However, TTA underestimates offspring counts in the case of co-first or co-last authorship, when first authors change the field of interest and in the case of alphabetical author lists or field-specific author ranking [63]. TTA like other name-dependent approaches faces the challenge of author disambiguation. This can

be mitigated by inclusion of initials, author affiliations and advanced computational algorithms [6, 64-67]. An ultimate solution is the assignment of unique and persistent author identifiers such as Orcid iDs [68] and their inclusion in publication records of bibliometric databases. Peer-reviewed articles were used to introduce the features of TTA as this form of publication represents the core of scientific production [69]. In the future, TTA may also be applied to other types such as preprints and patents.

Acknowledgments

The author thanks V. Demais, S. Eglen, J. Jouzel, J.M. Lehn, M. Muzet, V. Pallottini, J.L. Paluh, B. Poulain, H. Runz, J.P. Sauvage, D. Schulte, S. Silber and M. Slezak for discussions and helpful comments on previous versions of the manuscript.

References

1. Claxton LD. Scientific authorship: Part 2. History, recurring issues, practices, and guidelines. Mutation Research/Reviews in Mutation Research. 2005;589(1):31-45. doi: https://doi.org/10.1016/j.mrrev.2004.07.002.

2. Marusic A, Bosnjak L, Jeroncic A. A Systematic Review of Research on the Meaning, Ethics and Practices of Authorship across Scholarly Disciplines. Plos One. 2011;6(9). doi: 10.1371/journal.pone.0023477. PubMed PMID: WOS:000294802800004.

3. Bornmann L, Mutz R. Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. Journal of the Association for Information Science and Technology. 2015;66(11):2215-22. doi: 10.1002/asi.23329.

4. Clauset A, Larremore DB, Sinatra R. Data-driven predictions in the science of science. Science. 2017;355(6324):477-80. doi: 10.1126/science.aal4217. PubMed PMID: WOS:000393183100032.

5. Mukherjee S, Romero DM, Jones B, Uzzi B. The nearly universal link between the age of past knowledge and tomorrow's breakthroughs in science and technology: The hotspot. Science advances. 2017;3(4):e1601315. Epub 2017/04/26. doi: 10.1126/sciadv.1601315. PubMed PMID: 28439537; PubMed Central PMCID: PMCPMC5397134.

6. Zeng A, Shen ZS, Zhou JL, Wu JS, Fan Y, Wang YG, et al. The science of science: From the perspective of complex systems. Physics Reports-Review Section of Physics Letters. 2017;714:1-73. doi: 10.1016/j.physrep.2017.10.001. PubMed PMID: WOS:000418464100001.

7. Fortunato S, Bergstrom CT, Borner K, Evans JA, Helbing D, Milojevic S, et al. Science of science. Science. 2018;359(6379). Epub 2018/03/03. doi: 10.1126/science.aao0185. PubMed PMID: 29496846; PubMed Central PMCID: PMCPMC5949209.

8. Docampo D, Bessoule J-J. A new approach to the analysis and evaluation of the research output of countries and institutions. Scientometrics. 2019;119(2):1207-25. doi: 10.1007/s11192-019-03089-w.

9. Hardwicke TE, Serghiou S, Janiaud P, Danchev V, Crüwell S, Goodman SN, et al. Calibrating the Scientific Ecosystem Through Meta-Research. Annual Review of Statistics and Its Application. 2020;7(1):null. doi: 10.1146/annurev-statistics-031219-041104.

10. Melin G, Persson O. Studying research collaboration using co-authorships. Scientometrics. 1996;36(3):363-77. doi: 10.1007/bf02129600. PubMed PMID: WOS:A1996VG50500006.

11. Barabási AL, Jeong H, Néda Z, Ravasz E, Schubert A, Vicsek T. Evolution of the social network of scientific collaborations. Physica A: Statistical Mechanics and its Applications. 2002;311(3):590-614. doi: https://doi.org/10.1016/S0378-4371(02)00736-7.

12. Muller HM, Kenny EE, Sternberg PW. Textpresso: An ontology-based information retrieval and extraction system for biological literature. Plos Biology. 2004;2(11):1984-98. doi: 10.1371/journal.pbio.0020309. PubMed PMID: WOS:000225160300028.

13. Agarwal P, Searls DB. Can literature analysis identify innovation drivers in drug discovery? Nature Reviews Drug Discovery. 2009;8(11):865-78. doi: 10.1038/nrd2973. PubMed PMID: WOS:000271388200020.

14. Foster JG, Rzhetsky A, Evans JA. Tradition and Innovation in Scientists' Research Strategies. American Sociological Review. 2015;80(5):875-908. doi: 10.1177/0003122415601618. PubMed PMID: WOS:000362448400001.

15. Fonseca Bde P, Sampaio RB, Fonseca MV, Zicker F. Co-authorship network analysis in health research: method and potential use. Health research policy and systems. 2016;14(1):34. Epub 2016/05/04. doi: 10.1186/s12961-016-0104-5. PubMed PMID: 27138279; PubMed Central PMCID: PMCPMC4852432.

16. Stamatatos E. A Survey of Modern Authorship Attribution Methods. Journal of the American Society for Information Science and Technology. 2009;60(3):538-56. doi: 10.1002/asi.21001. PubMed PMID: WOS:000263935100009.

17. Borner K, Contractor N, Falk-Krzesinski HJ, Fiore SM, Hall KL, Keyton J, et al. A Multi-Level Systems Perspective for the Science of Team Science. Science translational medicine. 2010;2(49). doi: 10.1126/scitranslmed.3001399. PubMed PMID: WOS:000288437200001.

18. Cunningham H, Tablan V, Roberts A, Bontcheva K. Getting More Out of Biomedical Documents with GATE's Full Lifecycle Open Source Text Analytics. PLOS Computational Biology. 2013;9(2):e1002854. doi: 10.1371/journal.pcbi.1002854.

19. Milojevic S. Principles of scientific research team formation and evolution. Proceedings of the National Academy of Sciences of the United States of America. 2014;111(11):3984-9. Epub 2014/03/05. doi: 10.1073/pnas.1309723111. PubMed PMID: 24591626; PubMed Central PMCID: PMCPMC3964124.

20. Lee YN, Walsh JP, Wang J. Creativity in scientific teams: Unpacking novelty and impact. Research Policy. 2015;44(3):684-97. doi: 10.1016/j.respol.2014.10.007. PubMed PMID: WOS:000350837800010.

21. Xie Z, Ouyang Z, Li J. A geometric graph model for coauthorship networks. Journal of Informetrics. 2016;10(1):299-311. doi: https://doi.org/10.1016/j.joi.2016.02.001.

22. Miotto R, Wang F, Wang S, Jiang XQ, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. Briefings in Bioinformatics. 2018;19(6):1236-46. doi: 10.1093/bib/bbx044. PubMed PMID: WOS:000456689400011.

23. Zheng S, Dharssi S, Wu M, Li J, Lu Z. Text Mining for Drug Discovery. Methods in molecular biology (Clifton, NJ). 2019;1939:231-52. Epub 2019/03/09. doi: 10.1007/978-1-4939-9089-4_13. PubMed PMID: 30848465.

24. Chen Q, Lee K, Yan S, Kim S, Wei CH, Lu Z. BioConceptVec: Creating and evaluating literature-based biomedical concept embeddings on a large scale. PLoS

Comput Biol. 2020;16(4):e1007617. Epub 2020/04/24. doi: 10.1371/journal.pcbi.1007617. PubMed PMID: 32324731; PubMed Central PMCID: PMCPMC7237030.

25. Moral-Munoz JA, Herrera-Viedma E, Santisteban-Espejo A, Cobo MJ. Software tools for conducting bibliometric analysis in science: An up-to-date review. Profesional De La Informacion. 2020;29(1). doi: 10.3145/epi.2020.ene.03. PubMed PMID: WOS:000531809400004.

26. Venkatakrishnan AJ, Puranik A, Anand A, Zemmour D, Yao X, Wu X, et al. Knowledge synthesis of 100 million biomedical documents augments the deep expression profiling of coronavirus receptors. eLife. 2020;9:e58040. doi: 10.7554/eLife.58040.

27. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2019.

28. Dowle M. data.table: Extension of 'data.frame' 2019. Available from: https://CRAN.R-project.org/package=data.table.

29. Dinno A. dunn.test: Dunn's Test of Multiple Comparisons Using Rank Sums. 2017. Available from: https://CRAN.R-project.org/package=dunn.test.

30. Larsson J. eulerr: Area-Proportional Euler and Venn Diagrams with Ellipses. 2019. Available from: https://CRAN.R-project.org/package=eulerr.

31. Tang YH, M.; Li, W. ggfortify: Unified Interface to Visualize Statistical Result of Popular R Packages. The R Journal. 2016;8(2).

32. Wickham H. Ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag; 2016.

33. Slowikowski KS, A.; Hughes, S.; Lukauskas, S.; Irisson, J.O.; Kamvar, Z.N.; Thompson, R.; Dervieux, C.; Yutani, H.; Gramme, P. ggrepel: Automatically Position Non-Overlapping Text Labels with 'ggplot2'. 2018. Available from: https://CRAN.R-project.org/package=ggrepel.

34. Csardi GN, T. The igraph software package for complex network research. InterJournal. 2006;Complex Systems:1695.

35. Soetaert K. plot3D: Plotting Multi-Dimensional Data. 2017. Available from: https://CRAN.R-project.org/package=plot3D.

36. Newman ME. The structure of scientific collaboration networks. Proceedings of the National Academy of Sciences of the United States of America. 2001;98(2):404-9. Epub 2001/01/10. doi: 10.1073/pnas.021544898. PubMed PMID: 11149952; PubMed Central PMCID: PMCPMC14598.

37. Moroz LL. Aplysia. Current biology : CB. 2011;21(2):R60-1. Epub 2011/01/25. doi: 10.1016/j.cub.2010.11.028. PubMed PMID: 21256433; PubMed Central PMCID: PMCPMC4024469.

38. Hille F, Richter H, Wong SP, Bratovič M, Ressel S, Charpentier E. The Biology of CRISPR-Cas: Backward and Forward. Cell. 2018;172(6):1239-59. doi: https://doi.org/10.1016/j.cell.2017.11.032.

39. Garreta E, Kamm RD, Chuva de Sousa Lopes SM, Lancaster MA, Weiss R, Trepat X, et al. Rethinking organoid technology through bioengineering. Nature Materials. 2021;20(2):145-55. doi: 10.1038/s41563-020-00804-4.

40. Patke A, Young MW, Axelrod S. Molecular mechanisms and physiological importance of circadian rhythms. Nature Reviews Molecular Cell Biology. 2020;21(2):67-84. doi: 10.1038/s41580-019-0179-2.

41. Ioannidis JPA, Klavans R, Boyack KW. Thousands of scientists publish a paper every five days. Nature. 2018;561(7722):167-9. Epub 2018/09/14. doi: 10.1038/d41586-018-06185-8. PubMed PMID: 30209384.

42. Ioannidis JP, Boyack KW, Klavans R. Estimates of the continuously publishing core in the scientific workforce. PLoS One. 2014;9(7):e101698. Epub 2014/07/10. doi: 10.1371/journal.pone.0101698. PubMed PMID: 25007173.

43. Sauermann H, Haeussler C. Authorship and contribution disclosures. Science advances. 2017;3(11):e1700404. Epub 2017/11/21. doi: 10.1126/sciadv.1700404. PubMed PMID: 29152564; PubMed Central PMCID: PMCPMC5687853.

44. Sekara V, Deville P, Ahnert SE, Barabasi AL, Sinatra R, Lehmann S. The chaperone effect in scientific publishing. Proceedings of the National Academy of Sciences of the United States of America. 2018;115(50):12603-7. Epub 2018/12/12. doi: 10.1073/pnas.1800471115. PubMed PMID: 30530676; PubMed Central PMCID: PMCPMC6294962.

45. Li W, Aste T, Caccioli F, Livan G. Early coauthorship with top scientists predicts success in academic careers. Nature communications. 2019;10(1):5170. Epub 2019/11/16. doi: 10.1038/s41467-019-13130-4. PubMed PMID: 31729362; PubMed Central PMCID: PMCPMC6858367.

46. Wuchty S, Jones BF, Uzzi B. The increasing dominance of teams in production of knowledge. Science. 2007;316(5827):1036-9. Epub 2007/04/14. doi: 10.1126/science.1136099. PubMed PMID: 17431139.

47. Stallings J, Vance E, Yang J, Vannier MW, Liang J, Pang L, et al. Determining scientific impact using a collaboration index. Proceedings of the National Academy of Sciences of the United States of America. 2013;110(24):9680-5. Epub 2013/05/31.

doi: 10.1073/pnas.1220184110. PubMed PMID: 23720314; PubMed Central PMCID: PMCPMC3683734.

48. Coccia M, Wang L. Evolution and convergence of the patterns of international scientific collaboration. Proceedings of the National Academy of Sciences of the United States of America. 2016;113(8):2057-61. Epub 2016/02/03. doi: 10.1073/pnas.1510820113. PubMed PMID: 26831098; PubMed Central PMCID: PMCPMC4776471.

49. Parish AJ, Boyack KW, Ioannidis JPA. Dynamics of co-authorship and productivity across different fields of scientific research. PLoS One. 2018;13(1):e0189742. Epub 2018/01/11. doi: 10.1371/journal.pone.0189742. PubMed PMID: 29320509; PubMed Central PMCID: PMCPMC5761855.

50. Hirsch JE. An index to quantify an individual's scientific research output. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(46):16569-72. Epub 2005/11/09. doi: 10.1073/pnas.0507655102. PubMed PMID: 16275915; PubMed Central PMCID: PMCPMC1283832.

51. Bornmann L, Mutz R, Hug SE, Daniel H-D. A multilevel meta-analysis of studies reporting correlations between the h index and 37 different h index variants. Journal of Informetrics. 2011;5(3):346-59. doi: https://doi.org/10.1016/j.joi.2011.01.006.

52. Dinsmore A, Allen L, Dolby K. Alternative perspectives on impact: the potential of ALMs and altmetrics to inform funders about research impact. PLoS Biol. 2014;12(11):e1002003. Epub 2014/11/26. doi: 10.1371/journal.pbio.1002003. PubMed PMID: 25423184; PubMed Central PMCID: PMCPMC4244028.

53. Hutchins BI, Yuan X, Anderson JM, Santangelo GM. Relative Citation Ratio (RCR): A New Metric That Uses Citation Rates to Measure Influence at the Article Level.

PLoS Biol. 2016;14(9):e1002541. Epub 2016/09/07. doi: 10.1371/journal.pbio.1002541. PubMed PMID: 27599104; PubMed Central PMCID: PMCPMC5012559.

54. Ioannidis JP, Klavans R, Boyack KW. Multiple Citation Indicators and Their Composite across Scientific Disciplines. PLoS Biol. 2016;14(7):e1002501. Epub 2016/07/02. doi: 10.1371/journal.pbio.1002501. PubMed PMID: 27367269; PubMed Central PMCID: PMCPMC4930269.

55. Waltman L. A review of the literature on citation impact indicators. Journal of Informetrics. 2016;10(2):365-91. doi: 10.1016/j.joi.2016.02.007. PubMed PMID: WOS:000377413800004.

56. Fire M, Guestrin C. Over-optimization of academic publishing metrics: observing Goodhart's Law in action. GigaScience. 2019;8(6). Epub 2019/05/31. doi: 10.1093/gigascience/giz053. PubMed PMID: 31144712; PubMed Central PMCID: PMCPMC6541803.

57. McKiernan EC, Schimanski LA, Muñoz Nieves C, Matthias L, Niles MT, Alperin JP. Use of the Journal Impact Factor in academic review, promotion, and tenure evaluations. eLife. 2019;8:e47338. doi: 10.7554/eLife.47338.

58. Wislar JS, Flanagin A, Fontanarosa PB, Deangelis CD. Honorary and ghost authorship in high impact biomedical journals: a cross sectional survey. BMJ (Clinical research ed). 2011;343:d6128. Epub 2011/10/27. doi: 10.1136/bmj.d6128. PubMed PMID: 22028479; PubMed Central PMCID: PMCPMC3202014.

59. Al-Herz W, Haider H, Al-Bahhar M, Sadeq A. Honorary authorship in biomedical journals: how common is it and why does it exist? Journal of medical ethics.

2014;40(5):346-8. Epub 2013/08/21. doi: 10.1136/medethics-2012-101311. PubMed PMID: 23955369.

60. David SV, Hayden BY. Neurotree: a collaborative, graphical database of the academic genealogy of neuroscience. PLoS One. 2012;7(10):e46608. Epub 2012/10/17. doi: 10.1371/journal.pone.0046608. PubMed PMID: 23071595; PubMed Central PMCID: PMCPMC3465338.

61. Hirshman BR, Tang JA, Jones LA, Proudfoot JA, Carley KM, Marshall L, et al. Impact of medical academic genealogy on publication patterns: An analysis of the literature for surgical resection in brain tumor patients. Annals of neurology. 2016;79(2):169-77. Epub 2016/01/05. doi: 10.1002/ana.24569. PubMed PMID: 26727354.

62. Lienard JF, Achakulvisut T, Acuna DE, David SV. Intellectual synthesis in mentorship determines success in academic careers. Nature communications. 2018;9(1):4840. Epub 2018/11/30. doi: 10.1038/s41467-018-07034-y. PubMed PMID: 30482900; PubMed Central PMCID: PMCPMC6258699.

63. Waltman L. An empirical analysis of the use of alphabetical authorship in scientific publishing. Journal of Informetrics. 2012;6(4):700-11. doi: 10.1016/j.joi.2012.07.008. PubMed PMID: WOS:000308581700029.

64. Calero C, Buter R, Cabello Valdés C, Noyons E. How to identify research groups using publication analysis: an example in the field of nanotechnology. Scientometrics. 2006;66(2):365-76. doi: 10.1007/s11192-006-0026-z.

65. Perianes-Rodríguez A, Olmeda-Gómez C, Moya-Anegón F. Detecting, identifying and visualizing research groups in co-authorship networks. Scientometrics. 2010;82(2):307-19. doi: 10.1007/s11192-009-0040-z.

66. Milojevic S. Accuracy of simple, initials-based methods for author name disambiguation. Journal of Informetrics. 2013;7(4):767-73. doi: 10.1016/j.joi.2013.06.006. PubMed PMID: WOS:000327920400001.

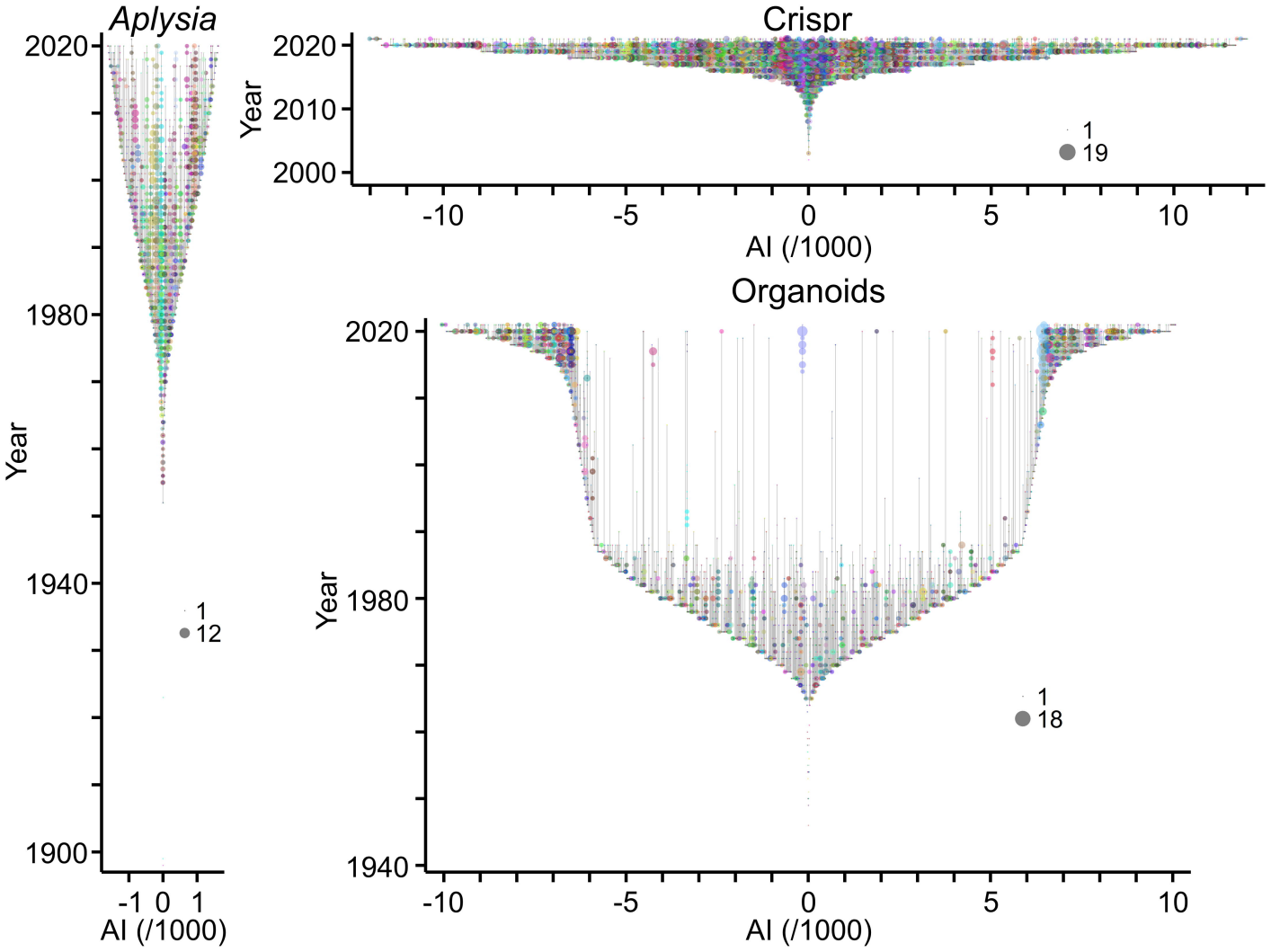
67. Lerchenmueller MJ, Sorenson O. Author Disambiguation in PubMed: Evidence on the Precision and Recall of Author-ity among NIH-Funded Scientists. PLoS One. 2016;11(7):e0158731. Epub 2016/07/02. doi: 10.1371/journal.pone.0158731. PubMed PMID: 27367860; PubMed Central PMCID: PMCPMC4930168.

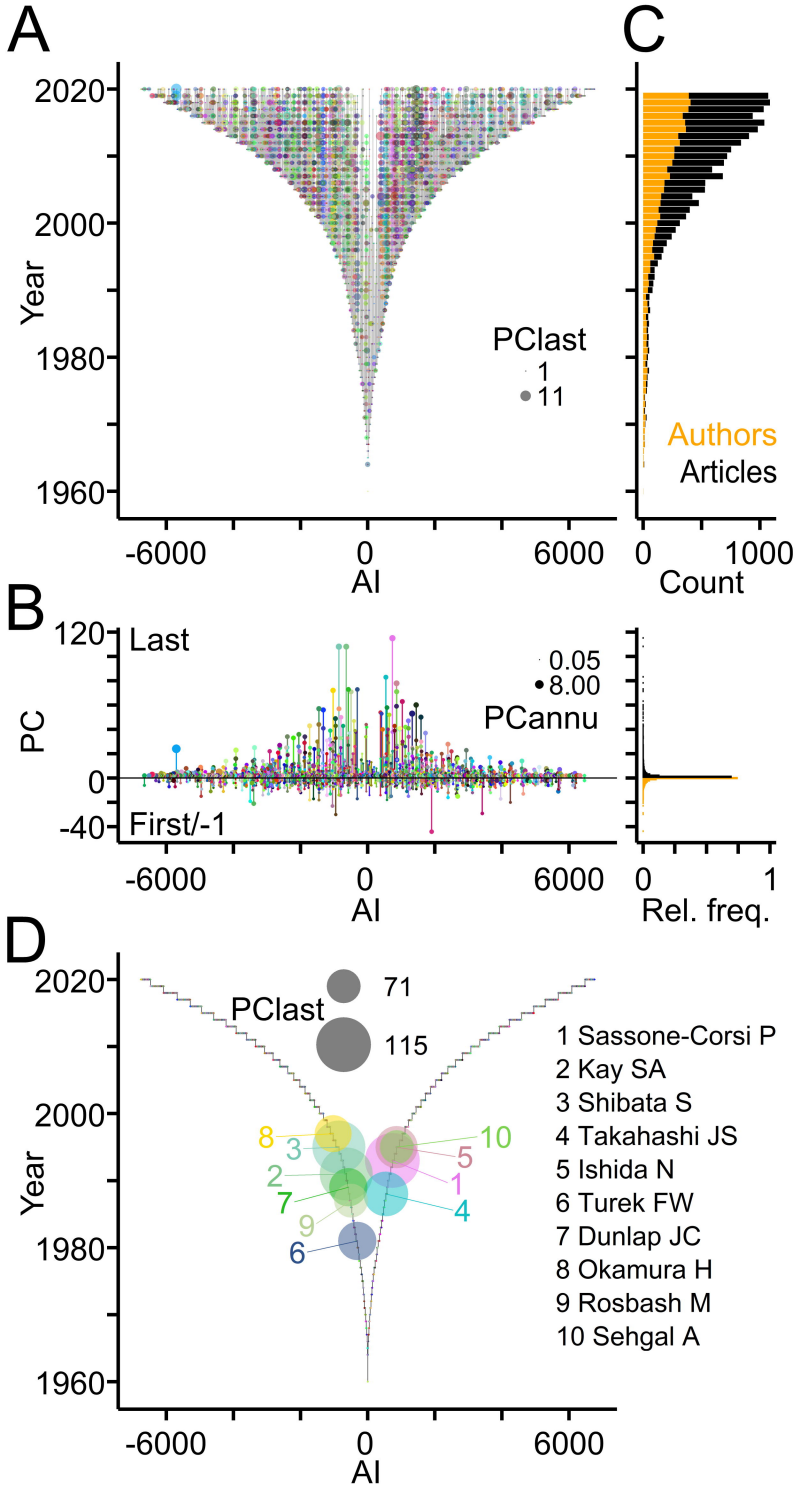
68. ORCID. Available from: https://orcid.org/.

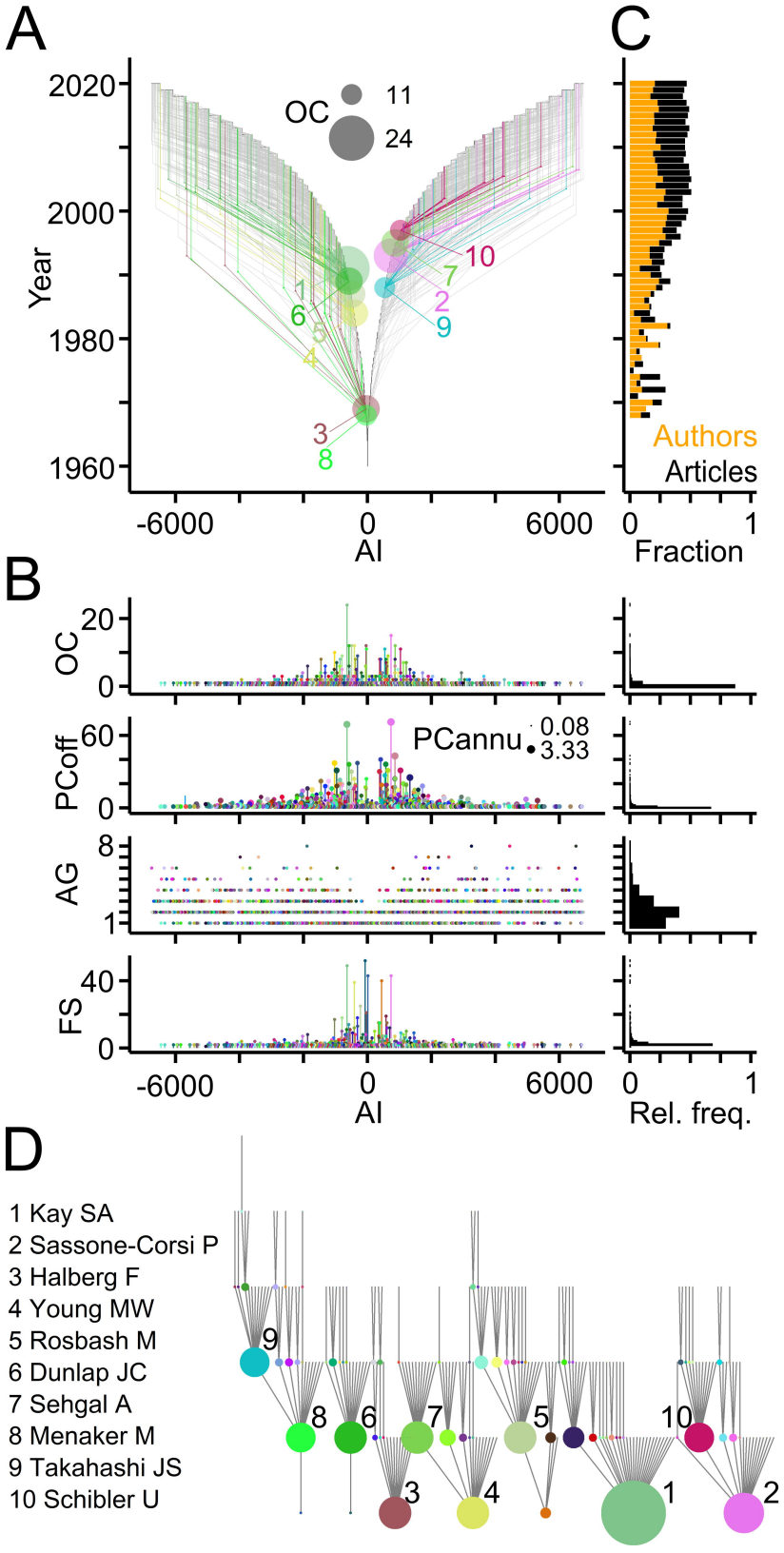
69. Tennant JP, Ross-Hellauer T. The limitations to our understanding of peer review. Research integrity and peer review. 2020;5:6. Epub 2020/05/06. doi: 10.1186/s41073-020-00092-1. PubMed PMID: 32368354; PubMed Central PMCID: PMCPMC7191707.

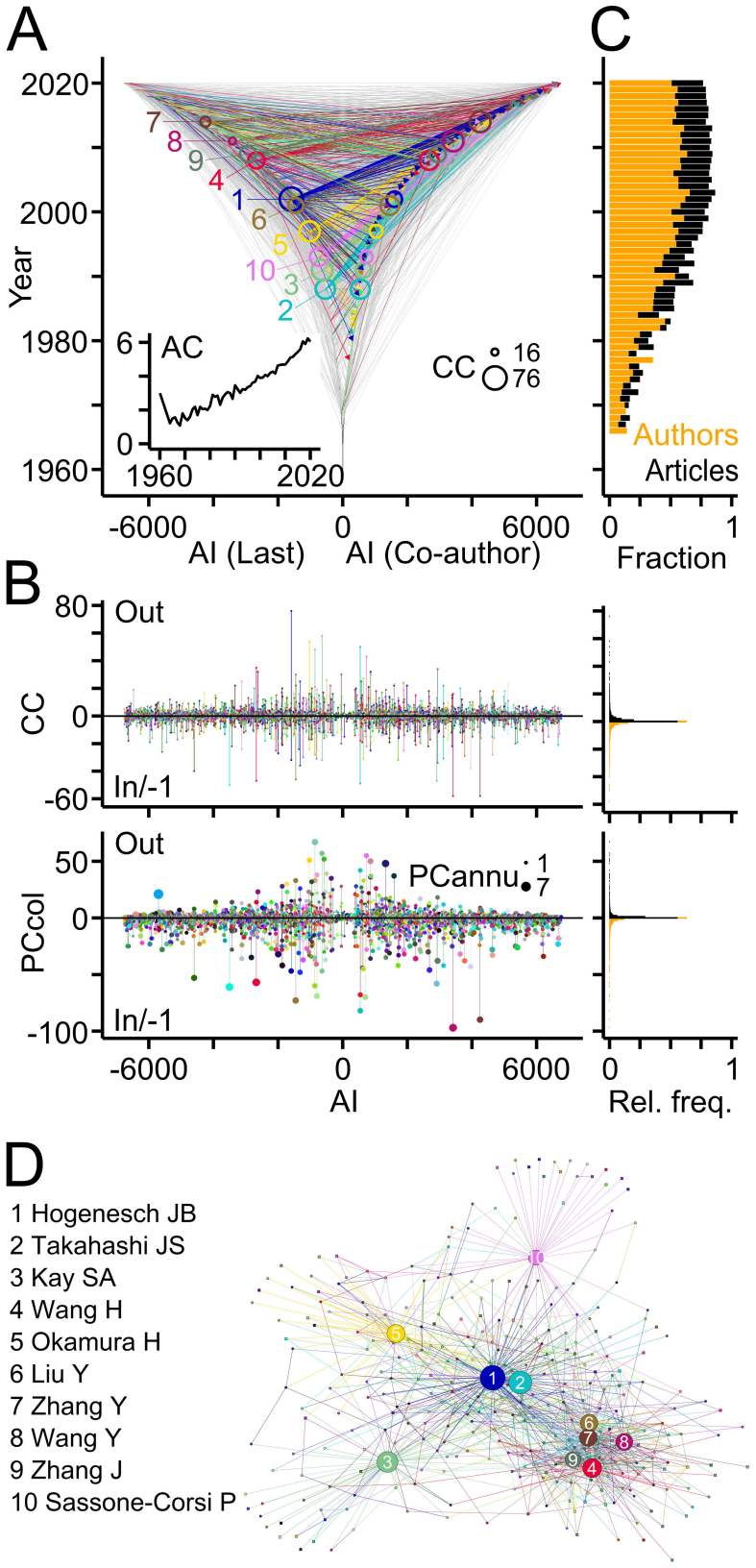
Supporting information

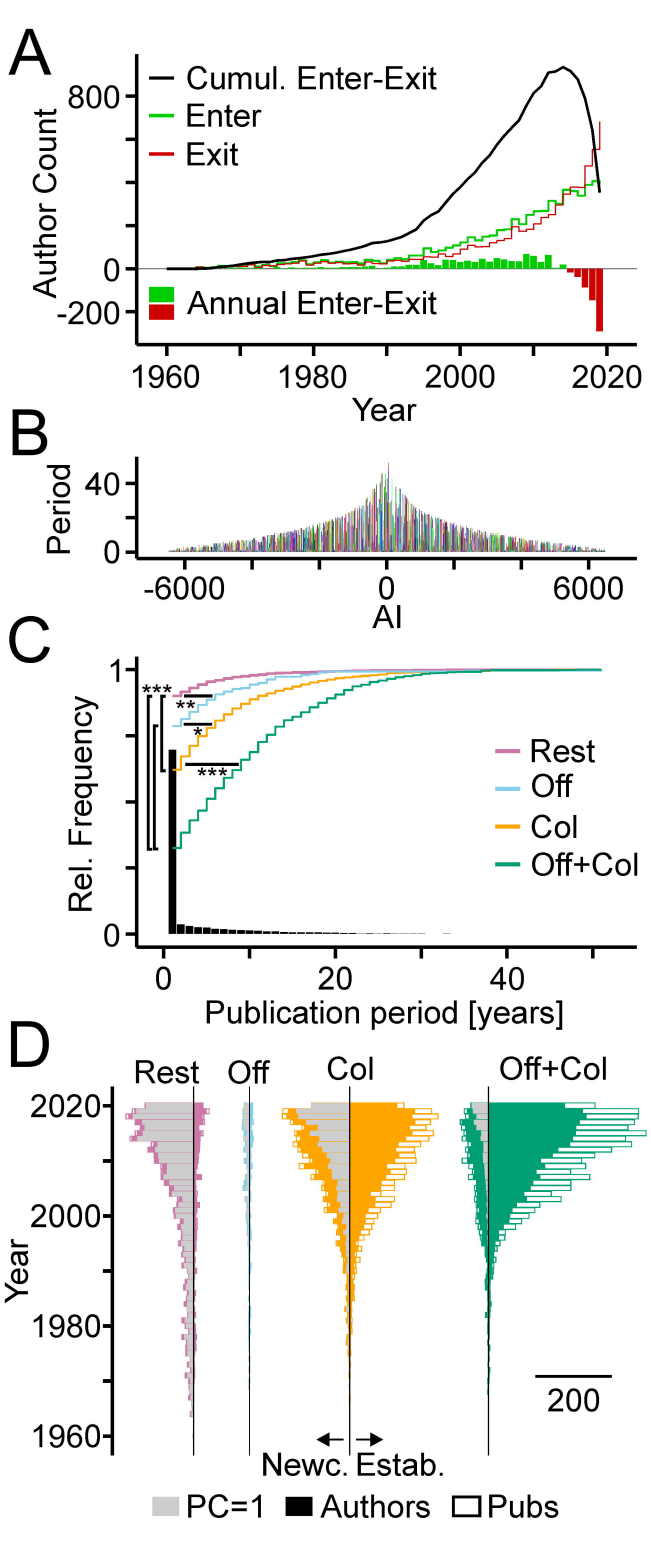
S1 File. TTA-derived results for the Clock field. Csv file summarizing TTA data for the Clock field using PubMed articles related to "circadian clock".

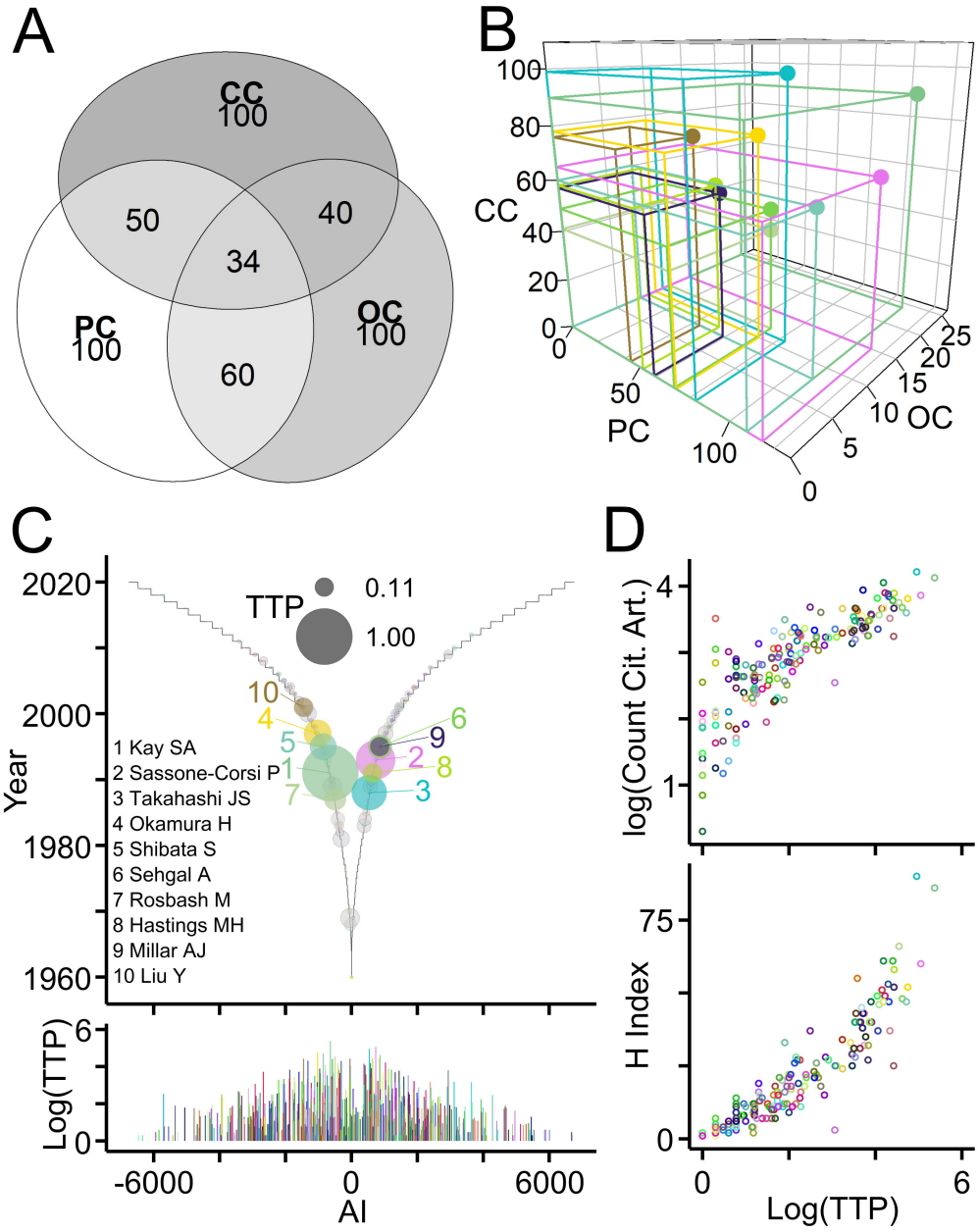

Aplysia
Year
2020
1980
1940
1900
1
12
-1 0 1
AI (/1000)
Crispr
Year
2020
2010
2000
1
19
-10
-5
0
5
10
AI (/1000)
Organoids
Year
2020
1980
1940
1
18
-10
-5
0
5
10
AI (/1000)

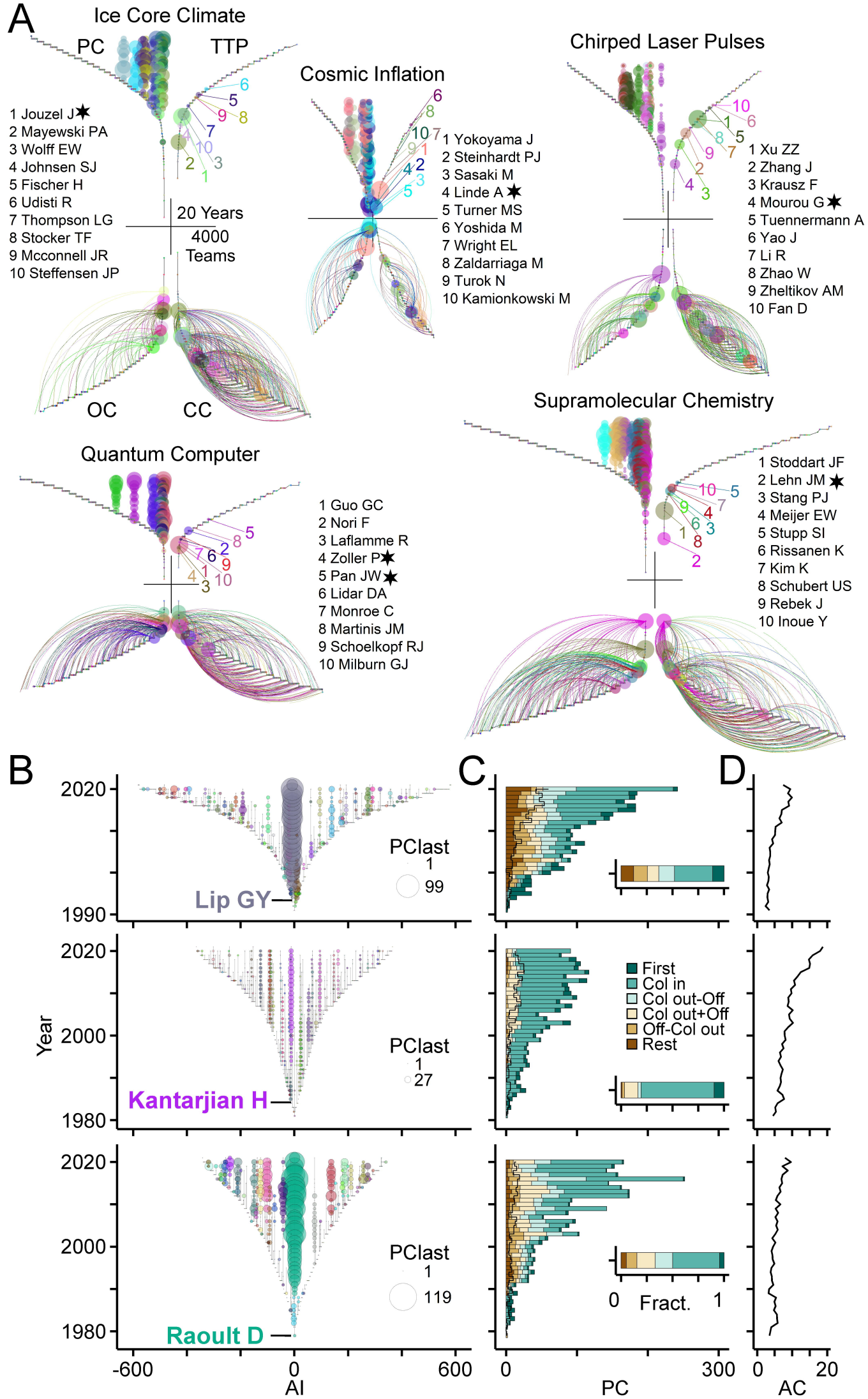

A
C
2020
2000
1980
1960
Year
PClast
1
11
-6000
0
6000
AI
Authors
Articles
0
1000
Count
B
120
Last
0.05
8.00
PCannu
PC
0
-40
First/-1
-6000
0
6000
AI
0
1
Rel. freq.
D
2020
2000
1980
1960
Year
PClast
71
115
8
3
2
7
9
6
10
5
1
4
1 Sassone-Corsi P
2 Kay SA
3 Shibata S
4 Takahashi JS
5 Ishida N
6 Turek FW
7 Dunlap JC
8 Okamura H
9 Rosbash M
10 Sehgal A
-6000
0
6000
AI


A
C
B
D
Year
2020
2000
1980
1960
OC
11
24
AI
-6000
0
6000
Authors
Articles
Fraction
0
1
OC
PCoff
AG
FS
PCannu
0.08
3.33
Rel. freq.
1 Kay SA
2 Sassone-Corsi P
3 Halberg F
4 Young MW
5 Rosbash M
6 Dunlap JC
7 Sehgal A
8 Menaker M
9 Takahashi JS
10 Schibler U


A
C
Year
2020
2000
1980
1960
AC
6
0
1960
2020
CC
16
76
-6000
0
6000
AI (Last)
AI (Co-author)
Authors
Articles
0
1
Fraction
B
CC
80
0
-60
Out
In/-1
PCcol
50
0
-100
PCannu
1
7
-6000
0
6000
AI
0
1
Rel. freq.
D
1 Hogenesch JB
2 Takahashi JS
3 Kay SA
4 Wang H
5 Okamura H
6 Liu Y
7 Zhang Y
8 Wang Y
9 Zhang J
10 Sassone-Corsi P


A
Cumul. Enter-Exit
Enter
Exit
Annual Enter-Exit
Author Count
800
0
-200
1960
1980
2000
2020
Year
B
Period
40
0
-6000
0
6000
AI
C
Rel. Frequency
1
0
***
**
*
***
Rest
Off
Col
Off+Col
0
20
40
Publication period [years]
D
Rest
Off
Col
Off+Col
Year
2020
2000
1980
1960
200
Newc. Estab.
PC=1
Authors
Pubs


A
CC
100
50
40
34
PC
100
OC
100
60
B
CC
PC
OC
C
TTP
0.11
1.00
Year
2020
2000
1980
1960
1 Kay SA
2 Sassone-Corsi P
3 Takahashi JS
4 Okamura H
5 Shibata S
6 Sehgal A
7 Rosbash M
8 Hastings MH
9 Millar AJ
10 Liu Y
Log(TTP)
AI
-6000
0
6000
D
log(Count Cit. Art.)
H Index
Log(TTP)


A
Ice Core Climate
PC
TTP
1 Jouzel J
2 Mayewski PA
3 Wolff EW
4 Johnsen SJ
5 Fischer H
6 Udisti R
7 Thompson LG
8 Stocker TF
9 Mcconnell JR
10 Steffensen JP
20 Years
4000
Teams
OC
CC
Cosmic Inflation
1 Yokoyama J
2 Steinhardt PJ
3 Sasaki M
4 Linde A
5 Turner MS
6 Yoshida M
7 Wright EL
8 Zaldarriaga M
9 Turok N
10 Kamionkowski M
Chirped Laser Pulses
1 Xu ZZ
2 Zhang J
3 Krausz F
4 Mourou G
5 Tuennermann A
6 Yao J
7 Li R
8 Zhao W
9 Zheltikov AM
10 Fan D
Quantum Computer
1 Guo GC
2 Nori F
3 Laflamme R
4 Zoller P
5 Pan JW
6 Lidar DA
7 Monroe C
8 Martinis JM
9 Schoelkopf RJ
10 Milburn GJ
Supramolecular Chemistry
1 Stoddart JF
2 Lehn JM
3 Stang PJ
4 Meijer EW
5 Stupp SI
6 Rissanen K
7 Kim K
8 Schubert US
9 Rebek J
10 Inoue Y
B
C
D
Year
2020
1990
2000
1980
PClast
1
99
Lip GY
27
Kantarjian H
119
Raoult D
-600
0
600
AI
First
Col in
Col out-Off
Col out+Off
Off-Col out
Rest
0
Fract.
1
300
PC
20
AC