

中图分类号: R95; G353.12 文献标志码: A 文章编号: 1006-4931(2023)22-0008-05
doi:10.3969/j.issn.1006-4931.2023.22.002



合成生物学高被引论文替代计量学指标和文献计量学指标与被引频次相关性分析*

潘黎姿¹, 龙俊羽², 袁艺², 陈娟¹, 欧阳昭连^{1Δ}

(1. 中国医学科学院医学信息研究所, 北京 100020; 2. 北京大学医学部, 北京 100191)

专家简介:潘黎姿,女,在读硕士研究生,大学本科毕业于北京大学,获生物医学英语专业学士学位,2022年至今就读于北京协和医学院情报学专业,攻读方向为医药科技情报研究。作为第一作者发表核心期刊论文2篇,参与发表中文核心期刊论文4篇和SCI论文2篇,参与编写多份医学前沿领域研究报告。



摘要:目的 探讨合成生物学高被引论文替代计量学指标和文献计量学指标与被引频次的相关性。方法 检索Web of Science数据库中合成生物学相关论文,检索时间为2023年6月21日,筛选高被引论文(被引频次排名前1%);同时,通过该数据库获取期刊引证报告(JCR)影响因子,与论文被引频次共同作为文献计量学指标,通过Altmetric官网获取Altmetric评分、Twitters等7个指标作为替代计量学指标;排除数据缺失比例超过20%的指标。分析合成生物学高被引论文发文量的年度变化趋势;分别采用多变量、单变量回归模型分析各指标与被引频次的相关性。结果 共纳入100篇高被引论文及6个指标[包括2个文献计量学指标(被引频次和JCR影响因子)及4个替代计量学指标(Altmetric评分、Twitters、Mendeley和Patents)]。1999年至2021年,合成生物学高被引论文发文量先上升后下降,集中发表于2011年至2015年(51篇)。多变量回归模型分析显示,Altmetric评分、Mendeley、Patents与被引频次呈显著正相关,JCR影响因子和Twitters与被引频次的相关性不显著。单变量回归模型分析显示,与被引频次的相关性除Twitters不显著外,其他变量均显著,其中JCR影响因子相关性

* 基金项目:国家自然科学基金[L1924064];北京协和医学院中央高校基本科研业务费项目[3332022041];中国医学科学院医学与健康科技创新工程项目[2021-I2M-1-056]。

第一作者:潘黎姿,女,在读硕士研究生,研究方向为医疗器械与药物战略情报,(电子信箱)pan.lizi@imicams.ac.cn。

Δ通信作者:欧阳昭连,女,博士,研究员,研究方向为医疗器械与药物战略情报,(电子信箱)ouyang.zhaolian@imicams.ac.cn。

- [19] 刘晓,张学博,陈大明,等. 2022年合成生物学发展态势[J]. 生命科学,2023,35(1):63-71.
- [20] 曹芹,旷苗,王晶,等. 我国“合成生物学”项目立项概况与实施管理建议[J]. 合成生物学,2020,1(4):495-502.
- [21] 张先恩. 中国合成生物学发展回顾与展望[J]. 中国科学(生命科学),2019,49(12):1543-1572.
- [22] AVIV-REUVEN S, ROSENFELD A. Publication patterns' changes due to the COVID-19 pandemic: a longitudinal and short-term scientometric analysis[J]. Scientometrics, 2021, 126(8):6761-6784.
- [23] CHENG S, LIU X, JIANG G, et al. Orthogonal Engineering of Biosynthetic Pathway for Efficient Production of Limonene in *Saccharomyces cerevisiae* [J]. ACS Synthetic Biology, 2019, 8(5):968-975.
- [24] YIN Y, LI Y, JIANG D, et al. De novo biosynthesis of liquiritin in *Saccharomyces cerevisiae*[J]. ACTA Pharmaceutica Sinica B, 2020,10(4):711-721.
- [25] TONG Y, LUO YF, GAO W. Biosynthesis of paclitaxel using synthetic biology [J]. Phytochemistry Reviews, 2022, 21(3): 863-877.
- [26] MUHAMMAD A, FENG X, RASOOL A, et al. Production of plant natural products through engineered *Yarrowia lipolytica*[J]. Biotechnology Advances, 2020,43:107555.
- [27] LIU X, CHENG J, ZHU X, et al. De Novo Biosynthesis of Multiple Pinocembrin Derivatives in *Saccharomyces cerevisiae*[J]. ACS Synthetic Biology,2020,9(11):3042-3051.
- [28] CHENG J, CHEN J, LIU X, et al. The origin and evolution of the diosgenin biosynthetic pathway in yam [J]. Plant Communications, 2021,2(1):1-9.
- [29] LYU Y, MARSAFARI M, KOFFAS M, et al. Optimizing Oleaginous Yeast Cell Factories for Flavonoids and Hydroxylated Flavonoids Biosynthesis [J]. ACS Synthetic Biology, 2019, 8(11):2514-2523.
- [30] TONG Y, LYU Y, XU S, et al. Optimum chalcone synthase for flavonoid biosynthesis in microorganisms [J]. Critical Reviews in Biotechnology, 2021,41(8):1194-1208.
- [31] ABRAMO G, D'ANGELO CA, DI COSTA F. How the Covid-19 crisis shaped research collaboration behaviour [J]. Scientometrics, 2022, 127(8):5053-5071.

(收稿日期:2023-07-07;修回日期:2023-08-10)

较弱, Mendeley 和 Patents 分别可解释被引频次 90.0% 和 85.6% 的变化。结论 替代计量学指标在一定程度上可反映论文的影响力; JCR 影响因子与论文影响力相关性弱或无关。

关键词: 合成生物学; 替代计量学; 文献计量学; 被引频次; 高被引论文; 相关性分析

Correlation Analysis of Altmetric Indicators and Bibliometric Indicators with Citation Frequency of Highly - Cited Papers in Synthetic Biology

PAN Lizi¹, LONG Junyu², YUAN Yi², CHEN Juan¹, OUYANG Zhaolian¹

(1. Institute of Medical Information, Chinese Academy of Medical Sciences, Beijing, China 100020; 2. Peking University Health Science Center, Beijing, China 100191)

Abstract: Objective To investigate the correlation of altmetric indicators and bibliometric indicators with citation frequency of highly - cited papers in synthetic biology. **Methods** Papers related to synthetic biology in the Web of Science were searched on June 21, 2023, and highly - cited papers (with the top 1% citation frequency) were selected. At the same time, the impact factor of Journal Citation Report (JCR) was obtained by the above database and taken together with citation frequency as the bibliometric indicators, seven indicators such as Altmetric score and Twitters were obtained by the Altmetric official website as the altmetric indicators, and the indicators with data missing ratio over 20% were excluded. The annual publication quantity of highly - cited papers in synthetic biology was analyzed, and the correlation of each indicator with citation frequency was analyzed by the multivariate and univariate regression models. **Results** A total of 100 highly - cited papers and six indicators were included, including two bibliometric indicators (citation frequency and JCR impact factor) and four altmetric indicators (Altmetric score, Twitters, Mendeley and Patents). From 1999 to 2021, the publication quantity of highly - cited papers in synthetic biology first increased and then decreased, with a large quantity published from 2011 to 2015 (51 papers). The multivariate regression model analysis showed that the Altmetric score, Mendeley and Patents were significantly positively correlated with citation frequency, while the JCR impact factor and Twitters were not significantly correlated with citation frequency. The univariate regression model analysis showed that all variables were significantly correlated with citation frequency except for Twitters, among which the JCR impact factor was weakly correlated with citation frequency, and Mendeley and Patents could explain 90.0% and 85.6% of the variance of citation frequency respectively. **Conclusion** Altmetric indicators can reflect the influence of the papers to a certain extent, but the JCR impact factor was weakly or not correlated with citation frequency.

Key words: synthetic biology; altmetrics; bibliometrics; citation frequency; highly - cited paper; correlation analysis

论文被引频次可反映其在研究领域的影响力,是目前用于评判个人研究成果重要的传统指标。期刊引证报告(JCR)影响因子较高的期刊,其论文被引频次可能更高^[1],但同一期刊的论文被引频次可能差异较大^[2]。随着互联网的快速发展,越来越多的科研成果以电子形式在网络发布, Twitter, YouTube, Facebook, Blog 等社交媒体和主流媒体平台逐渐成为传播研究成果的常见工具,加强了研究人员之间及其与社会的联系,且可能影响论文被引频次^[3]。PRIEM 等^[4]提出了以 Altmetric (替代计量学)指标来衡量论文基于社交网络的影响力。Altmetric 网站是 2011 年由英国人 ADIE 开发的一款研究工具,通过提取单篇论文在不同社交网络平台和在线媒体提及的次数,以综合计算论文网络影响力。Altmetric 网站提供文章在十二大数据源中的表现数据,包括公共政策文件、博客文章、主流媒体报道、Web of Science 引文数据、在线文献管理软件、同行评审出版平台、学术社交论坛(Research highlights)、社交媒体(Facebook, Twitter, LinkedIn, Google+)等。合成生物学旨在通过挖掘基因功能元件和模块,对底盘细胞代谢调控网络进行基因设计、修改或补充,使活细胞产物满足人类的需求,是生命科学、工程学、信息学等多种学科融合而成的一门新兴学科^[5]。本研究中探讨了合成生

物学领域高被引论文替代计量学指标和文献计量学指标与被引频次的相关性。现报道如下。

1 资料与方法

1.1 文献检索

采用主题检索方式,在 Web of Science 数据库的科学引文索引扩展版(SCIE)核心合集中以摘要、关键词和标题字段检索合成生物学相关论文,检索式为(TS = ("synthetic biolog*") OR SO = (ACS SYNTHETIC BIOLOGY));在检索结果中选择“article”类型的论文,且对其他字段不作限制。将检索结果按被引频次由高到低排序,以排名前 1% 的论文作为高被引论文,统计其标题、作者、发表时间、发表期刊和被引频次。检索时间为 2023 年 6 月 21 日,对论文的发表时间不作限定。

1.2 数据获取与预处理

数据获取:通过 Web of Science 数据库获取高被引论文被引频次和 JCR 影响因子作为文献计量学指标。通过 Altmetric 网站获取替代计量学指标。1) Altmetric 评分,为 Altmetric 网站根据论文在不同公共平台的曝光情况加权计算获得的综合评分,通常用以衡量论文在公共平台的综合表现;2) Twitters,为论文在 Twitter 的分享次数;3) Facebook Pages,为论文在 Facebook 的分享次数;4) Mendeley,为论文用 Mendeley 软件保存的次数,该

软件可提供一系列文献管理功能,能有效地对文献进行组织、标注和存储管理;5) Patents, 为论文被专利引用的次数,专利是指对产品、方法或其改进所提出的新的技术方案,专利申请材料中需对方案进行详细描述,必要的原理阐述部分需引用文章等依据;6) Blogs, 为论文被 Blog 引用的次数;7) Weibo User, 为论文在微博(Weibo)中被引用的次数。各指标数据获取时间与论文检索时间一致。

数据预处理:将所有数据汇总,查看缺失情况,排除缺失比例超 20% 的指标;其余指标用平均值填补缺失值。数据填补完整后,将所有指标值进行归一化处理,即

$$x_{i,new} = \frac{x_i - \min(x_i)}{\max(x_i) - \min(x_i)}$$

其中, i 表示上述 9 个指标中的第 i 个指标, x_i 表示指标原值, $x_{i,new}$ 表示归一化后的指标值, $\min(x_i)$ 和 $\max(x_i)$ 分别表示 x_i 的最小值和最大值。

1.3 数据分析

采用 R Studio 软件分析高被引论文的年度发表情况,分别采用多变量、单变量回归模型分析文献计量学指标和替代计量学指标与被引频次的相关性。

2 结果

2.1 高被引论文概况

共获得 9 784 篇论文,以被引频次排名前 100 的论文为高被引论文,其中有 2 篇会议论文(被引频次排名分别为第 23 和第 54)无法获取替代计量学指标而不纳入,故选取排名为第 101 和第 102 的论文递补。

指标数据缺失情况见表 1。排除 Facebook Pages, Blogs, Weibo User 3 个指标,纳入 6 个指标(见表 2),其中包括文献计量学指标 2 个,替代计量学指标 4 个。将缺失数据填补并将所有数据归一化处理后,形成最终的数据集。

表 1 指标数据缺失情况 (%)

Tab. 1 Data missing of each indicator (%)

指标	数据缺失	指标	数据缺失
被引频次	0	Mendeley	0
JCR 影响因子	0	Patents	10.25
Altmetric 评分	0	Blogs	38.46
Twitters	18.80	Weibo User	96.58
Facebook Pages	61.53		

表 2 纳入指标的描述性统计

Tab. 2 Descriptive statistics of included indicators

指标	最小值	最大值	中位数	平均值
被引频次	227	6 195	302	431
JCR 影响因子	0.66	13	3	5
Altmetric 评分	2	1 209	30	109
Twitters	1	1 078	23	64
Mendeley	13	9 472	653	833
Patents	1	939	15	29

1999 年至 2021 年,高被引论文的发表趋势为先上升后下降,集中发表于 2011 年至 2015 年(51 篇);2015 年后,发文数量明显下降,可能是由于论文从发表到积累一定量的被引频次需要时间。详见图 1。

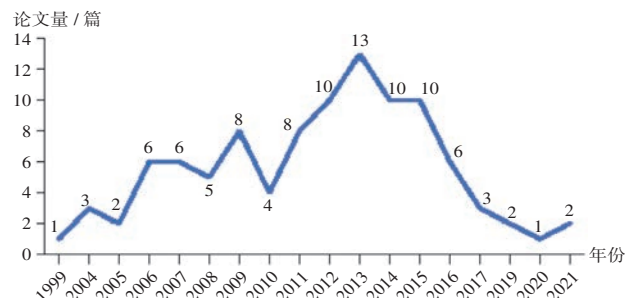


图 1 100 篇合成生物学高被引论文年度发表情况

Fig. 1 Annual publication of 100 highly-cited papers in synthetic biology

2.2 回归模型分析

多变量回归模型:以被引频次为因变量,文献计量学指标和替代计量学指标为自变量,共拟合 2 个多变量回归模型。模型 I 探究文献计量学指标和替代计量学指标对被引频次的总体影响。由于 Altmetric 评分可反映替代计量学指标的综合情况,故仅以 Altmetric 评分和 JCR 影响因子为自变量。结果显示,Altmetric 评分与被引频次呈显著正相关,JCR 影响因子与被引频次相关性不显著;该模型仅可解释被引频次 8.8% 的变化。模型 II 探究不同细化指标与被引频次之间的关系,故将 Altmetric 评分替换为 Twitters, Mendeley, Patents。结果显示, Mendeley 和 Patents 与被引频次均呈显著正相关, JCR 影响因子和 Twitters 与被引频次相关性不显著; Twitters 的影响力低于 Mendeley 和 Patents;该模型可解释被引频次 92.9% 的变化。详见表 3(其中 95% CI 为 95% 置信区间,表 4 同)。

表 3 多变量回归模型分析结果

Tab. 3 Results of multivariate regression model analysis

模型	自变量	β (95% CI)	R^2
模型 I	JCR 影响因子	0.017(-0.046, 0.081)	0.088
	Altmetric 评分	0.155(0.038, 0.273)**	
模型 II	JCR 影响因子	0.006(-0.012, 0.023)	0.929
	Twitters	-0.020(-0.061, 0.021)	
	Mendeley	0.601(0.488, 0.716)***	
	Patents	0.389(0.272, 0.506)***	

注:*, **, *** 分别表示在 10%, 5%, 1% 水平上显著。表 4 同。

Note: *, ** and *** indicate that the differences were significant at the 10%, 5% and 1% levels respectively (for Tab. 3 - 4).

单变量回归模型:对 5 个自变量分别进行拟合(见表 4)。结果显示,仅 Twitters 的拟合结果不显著。拟合结果显著的 4 个模型中, Mendeley 和 Patents 分别解释了被引频次 90.0% 和 85.6% 的变化, JCR 影响因子和 Altmetric 评分仅分别解释了被引频次 0.1% 和 8.6% 的变化。

表4 单变量回归模型分析结果

Tab. 4 Results of univariate regression model analysis

自变量	$\beta(95\% CI)$	R^2
JCR 影响因子	0.058(8.911 × 10 ⁻⁵ , 0.115)*	0.001
Altmetric 评分	0.171(0.068, 0.274)**	0.086
Twitters	0.013(-0.126, 0.152)	< 0.001
Mendeley	0.940(0.882, 0.998)***	0.900
Patents	0.951(0.879, 1.022)***	0.856

3 讨论

本研究中,多变量回归的2个模型分析结果显示JCR影响因子与被引频次无显著相关性,单变量回归模型分析结果显示JCR影响因子与被引频次呈弱相关,提示合成生物学高被引论文的被引频次与期刊的JCR影响因子相关性弱或无关。这与文献[6-7]的研究结果一致。提示高JCR影响因子的期刊对论文被引频次的贡献可能较低,提示学者应专注提高论文水平,而非一味追求高JCR影响因子。多变量、单变量回归模型分析结果均显示替代计量学指标(Altmetric评分)与被引频次显著相关,且国内外多位学者用不同的替代计量学指标和统计学方法得出公共平台对论文引用有正向影响的结论^[8-12]。提示在互联网高速发展的时代,论文影响力的传统评价体系已发生改变,互联网的及时性可提高论文的传播速度和范围。科研人员在发表科研成果的同时,应充分利用公共社交媒体平台,及时发布最新的研究进展,增加公众和其他学者对研究成果的了解和关注,从而提高论文的被引频次和影响力。

本研究中, Twitters 与被引频次无显著相关性, Mendeley 和 Patents 相关性均较高。有研究发现, Twitters 对论文被引频次的影响程度在不同学科中不同,其中对普外科^[13]、泌尿外科^[14]、神经外科^[15]学术论文被引频次的影响程度较低,对心血管外科^[16]、整形外科^[17]、儿科^[18]学术论文的影响程度较高。分析原因,心血管外科、整形外科和儿科患者群体大、公众关注度高或主题与日常生活贴近,更易在 Twitter 平台传播,故 Twitters 与被引频次相关性高。合成生物学是一门新兴学科,发展历史短,研究内容较前沿,难以受到大众的关注,故在 Twitter 平台传播效果欠佳。Mendeley 作为一款文献管理软件,其使用者多为专业学者和研究人员;专利申请因具有创新性和新颖性,需紧跟领域最新理论进展来支撑技术创新。故上述两类使用人员更加关注行业前沿,在合成生物学领域表现出与论文被引频次显著的相关性。本研究的不足之处在于,仅对合成生物学领域的高被引论文进行研究,在研究样本的选择上有一定偏倚,故对不同学科及不同引用水平的论文可能产生不同的结论,未来应扩大研究范围,针对更多学科的全范围论文进行建模,同时纳入更多指标,获得更普遍适用的结论。

综上所述,合成生物学高被引论文的被引频次与期刊JCR影响因子相关性小或无关,替代计量学指标(除 Twitters 外)在一定程度上可反映论文的影响力。未来可利用替代计量学指标预测论文的影响力或将其融入现有文献评价系统中,使研究成果影响力的评价方法更符合当代互联网时代的发展趋势。

参考文献

- [1] 梁国强,侯海燕,任佩丽,等. 高质量论文使用次数与被引次数相关性的特征分析[J]. 情报杂志,2018,37(4):147-153.
- [2] PATEL RB, VADUGANATHAN M, MOSARLA RC, et al. Open Access Publishing and Subsequent Citations Among Articles in Major Cardiovascular Journals[J]. Am J Med, 2019, 132(9):1103-1105.
- [3] 吴朋民,陈挺,王小梅. Altmetrics 与引文指标相关性研究[J]. 数据分析与知识发现,2018,2(6):58-69.
- [4] PRIEM J, GROTH P, TARABORELLI D, et al. The altmetrics collection[J]. PLoS One, 2012, 7(11):e48753.
- [5] 王浩琦,高豪,信丰学. “十四五”背景下合成生物学产业发展趋势分析[J]. 生物学杂志,2023,40(3):1-5.
- [6] TSAY MY. The relationship between journal use in a medical library and citation use[J]. Bull Med Libr Assoc, 1998, 86(1):31-39.
- [7] SILVA DO, TABORDA B, PAZZINATTO MF, et al. The Altmetric Score Has a Stronger Relationship With Article Citations Than Journal Impact Factor and Open Access Status: A Cross-sectional Analysis of 4022 Sport Sciences Articles[J]. J Orthop Sports Phys Ther, 2021, 51(11):536-541.
- [8] 付大军,何海燕. 跨学科文献的多类选择性计量指标的定量比较研究[J]. 情报学报,2015,34(6):600-607.
- [9] 匡登辉. 高被引论文的论文级计量分析[J]. 情报杂志,2016,35(5):149-154.
- [10] 舒非,HAUSTEIN S,全薇. 推特(Twitter)对中国论文的国际关注度影响研究[J]. 图书馆论坛,2017,37(6):55-60.
- [11] NIP I, FENG H. Examining correlation of altmetric score and citation number in dermatology journal articles[J]. Journal of Dermatological Treatment, 2022, 33(1):297-299.
- [12] LOOMBA RS, VILLARREAL EG, PATEL RD, et al. Altmetric score, reads, and citations in paediatric cardiology: do they correlate and what do they mean? [J]. Cardiology Young, 2022, 20:1-6.
- [13] MULLINS CH, BOYD CJ, COREY BL. Examining the Correlation Between Altmetric Score and Citations in the General Surgery Literature[J]. J Surg Res, 2020(248):159-164.
- [14] NOCERA AP, BOYD CJ, BOUDREAU H, et al. Examining the Correlation Between Altmetric Score and Citations in the Urology Literature[J]. Urology, 2019(134):45-50.
- [15] WANG J, ALOTAIBI NM, IBRAHIM GM, et al. The Spectrum of Altmetrics in Neurosurgery: The Top 100 "Trending" Articles in Neurosurgical Journals[J]. World Neurosurg, 2017(103):