

基于论文分析的中美 CRISPR/Cas 领域 对比研究

孙劲松, 郑彦宁, 袁 芳

(中国科学技术信息研究所, 北京 100038)

摘要: 成簇规律间隔短回文重复序列 / 成簇规律间隔短回文重复序列关联蛋白 (CRISPR/Cas) 技术是基因编辑领域的研究热点, 目前广泛应用于生命科学研究、疾病模型构建、药物研发以及农业生产等领域。本文基于 2008—2018 年的 CRISPR/Cas 科技论文数据, 利用文献计量方法, 从论文逐年发表数量、主要国家排名、合作国家分布、主要研究机构以及高被引论文等角度对中美 CRISPR/Cas 技术领域的基础研究进展进行了比较分析, 旨在为政府、科研机构以及相关企业提供参考。

关键词: 美国; 中国; 论文分析; CRISPR/Cas 技术

中图分类号: G327.712 **文献标识码:** A **DOI:** 10.3772/j.issn.1009-8623.2019.01.008

随着生命科学研究进入基因组时代, 人们改写生物体内遗传序列信息的需求越来越迫切。基因编辑技术是一项对生物体内源基因进行精准定点修饰的技术, 是目前生物科学研究领域应用最广泛的技术^[1]。它通过对目标基因进行编辑, 进而实现对特定 DNA 片段的敲除和加入等。目前基因编辑主要包括应用锌指核酸内切酶技术 (Zinc-Finger Nucleases, ZFN)^[2]、类转录激活因子效应物核酸酶技术 (Transcription Activator-Like Effector Nuclease, TALEN)^[3] 和成簇规律间隔短回文重复序列 / 成簇规律间隔短回文重复序列关联蛋白技术 (Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-associated proteins, CRISPR/Cas)^[4]。其中, CRISPR/Cas 技术是近几年兴起并迅速发展的一项基因编辑技术, 其具备设计简单、操作方便、效率高、成本低且可同时进行多点编辑的优点^[5], 因此成为了最具广阔前景的基因编辑技术。目前 CRISPR/Cas 技术广泛应用于生命科学研究、疾病模型构建、药物研发以及农业生产等领域^[6]。一般而言, 科技论文是科学研究成果的重要载体和科研产出的主要形式, 对论文进行分析, 有助于了解学科领域的基础研究水平^[7]。因此, 本研究从论文分析的角度出发, 对中美两国在 CRISPR/Cas 技术领域的基础研究现状进行对比, 以期政府、科研机构以及相关企业的科技计划和技术研发提供一定的参考。

随着生命科学研究进入基因组时代, 人们改写生物体内遗传序列信息的需求越来越迫切。基因编辑技术是一项对生物体内源基因进行精准定点修饰的技术, 是目前生物科学研究领域应用最广泛的技术^[1]。它通过对目标基因进行编辑, 进而实现对特定 DNA 片段的敲除和加入等。目前基因编辑主要包括应用锌指核酸内切酶技术 (Zinc-Finger Nucleases, ZFN)^[2]、类转录激活因子效应物核酸酶技术 (Transcription Activator-Like Effector Nuclease, TALEN)^[3] 和成簇规律间隔短回文重复序列 / 成簇规律间隔短回文重复序列关联蛋白技术 (Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-associated proteins, CRISPR/Cas)^[4]。其中, CRISPR/Cas 技术是近几年兴起并迅速发展的一项基因编辑技术, 其具备设计简单、操作方便、效率高、成本低且可同时进行多点编辑的优点^[5], 因此成为了最具广阔前景的基因编辑技术。目前 CRISPR/Cas 技术广泛应用于生命科学研究、疾病模型构建、药物研发以及农业生产等领域^[6]。一般而言, 科技论文是科学研究成果的重要载体和科研产出的主要形式, 对论文进行分析, 有助于了解学科领域的基础研究水平^[7]。因此, 本研究从论文分析的角度出发, 对中美两国在 CRISPR/Cas 技术领域的基础研究现状进行对比, 以期政府、科研机构以及相关企业的科技计划和技术研发提供一定的参考。

1 数据来源

本研究选取美国汤森路透公司开发的 Web of Science 核心合集集中的 Science Citation Index Expanded (SCI-Expanded) 论文数据库。该数据库涵盖了自然科学、工程技术、社会科学、艺术与人文等诸多领域内的 8 500 多种学术期刊^[8]。根据研究主题, 设定检索式如下: (“clustered regularly interspaced short palindromic repeats”) OR (“clustered regularly interspaced short palindromic repeat”) OR

第一作者简介: 孙劲松 (1994—), 女, 在读硕士研究生, 主要研究方向为前沿领域分析与专利分析。

收稿日期: 2018-12-20

(“CRISPR\$”) OR (“CRISPER” and “Cas”) OR (“CRISPER” and “Cas9”)。将时间跨度设定为 2008—2018 年, 检索结果精炼为 Article, 共得到 7 729 篇论文, 其中美国发表论文 3 820 篇, 中国发表论文 1 819 篇, 德国发表论文 686 篇, 日本发表论文 676 篇, 英国发表论文 518 篇, 检索时间为 2018 年 12 月 10 日。

2 结果分析

2.1 中美 CRISPR/Cas 论文逐年发表情况

通过对全球及中美 2008—2018 年发表的 CRISPR/Cas 论文数量进行统计, 可以直观地看出该领域目前的研究水平、发展规模和研究速度^[9]。如图 1 所示, 全球 CRISPR/Cas 论文发表数量近 10 年呈逐步上升趋势, 说明全球对该领域的基础研究越来越重视。2008—2012 年间, 全球 CRISPR/Cas 论文发表数量增长缓慢。从 2012 年开始, 全球 CRISPR/Cas 论文数量增幅明显, 基本呈指数型趋

势上升。2018 年, 全球 CRISPR/Cas 论文数量达到峰值 2 403 篇。

中美近 10 年 CRISPR/Cas 领域年发文量的变化趋势与全球变化趋势基本一致。2008—2018 年间, 美国 CRISPR/Cas 论文数量一直多于中国。2008 年美国发表 CRISPR/Cas 论文 16 篇, 而中国在 2008—2012 年间的年发文量则一直只有个位数。2012—2017 年间, 中国和美国的 CRISPR/Cas 论文数量差距逐渐增大, 表明美国在该领域的研究发展速度加快。2013 年, 中国发表 CRISPR/Cas 论文数量仅为 14 篇, 还未达到美国 2008 年的数量水平, 可见中国在该领域的基础研究发展较慢。2017 年, 美国发表 CRISPR/Cas 论文 1 023 篇, 占全球 CRISPR/Cas 论文总量的一半; 中国发表 CRISPR/Cas 论文 514 篇, 是美国论文数量的一半。2018 年, 美国 CRISPR/Cas 论文数量达到峰值 1 125 篇, 中国 CRISPR/Cas 论文数量达到峰值 704 篇, 中美之间数量差距减小。

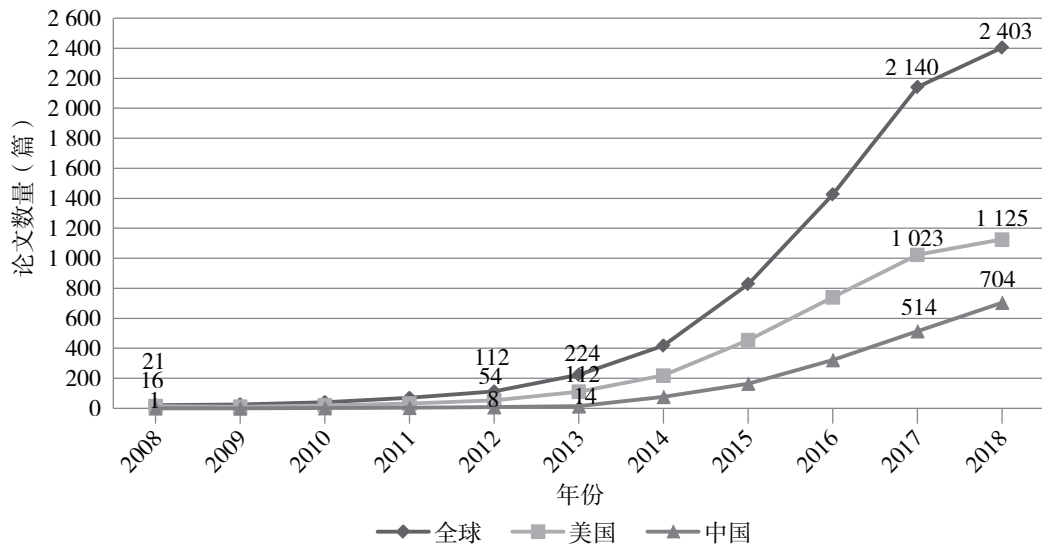


图 1 SCI 收录的 CRISPR/Cas 技术领域论文的发表数量变化

2.2 中美 CRISPR/Cas 论文世界排名情况

通过统计 CRISPR/Cas 技术的国家或地区论文产出, 能够呈现全球范围的基础研究布局^[10]。近 10 年共有 109 个国家或地区在 CRISPR/Cas 领域发表了相关论文。从图 2 可见, 在发表论文数量最多的前 15 位国家中, 美国、中国、德国、日本和英国分别位居前 5 位。美国发表 CRISPR/Cas 论文数量最多, 共 3 820 篇; 中国仅次于美国,

发表 CRISPR/Cas 论文 1 819 篇; 其他国家发表的 CRISPR/Cas 论文数量均不足 1 000 篇。可见, 中美两国的 CRISPR/Cas 论文数量远远超过了其他国家, 处于世界领先地位。但美国的 CRISPR/Cas 论文数量是中国的 2 倍, 表明中国与世界领军国家美国之间还存在较大的差距。此外, 中国是 CRISPR/Cas 论文数量排名前 15 位国家中唯一的发展中国家, 论文数量远超过除美国以外的其他发达国家, 说明中国

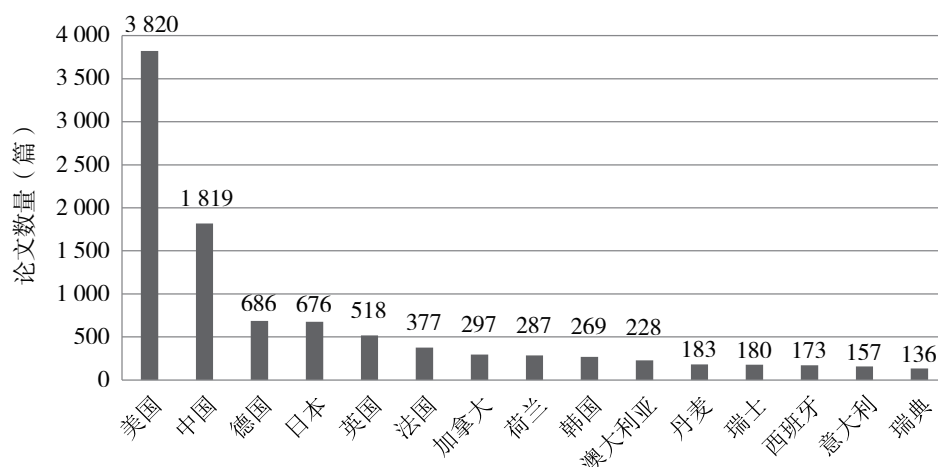


图2 SCI收录的CRISPR/Cas技术领域论文数量最多的前15位国家

在该领域的基础研究取得了相当不错的成果。

2.3 中美 CRISPR/Cas 论文合作国家分布情况

通过对某一领域论文的合作国家分布进行分析,可以了解不同国家在该领域研究的合作密切程度^[11]。从表1可见,中美两国在CRISPR/Cas技术领域的合作伙伴不尽相同。美国的前5位合

作国家分别是中国、德国、英国、日本和法国,而中国的前5位合作国家分别是美国、英国、日本、丹麦和加拿大。中美两国近10年合作发表的CRISPR/Cas论文数量最多,达到582篇,而中国与其他国家合作发表的CRISPR/Cas论文数量均不足100篇,可见中美两国在该领域的研究合作最

表1 中美在CRISPR/Cas技术领域的前10位合作国家

序号	美国	合作论文数量	中国	合作论文数量
1	中国	582	美国	582
2	德国	192	英国	59
3	英国	182	日本	33
4	日本	147	丹麦	30
5	法国	134	加拿大	29
6	加拿大	121	澳大利亚	27
7	荷兰	87	德国	25
8	韩国	82	新加坡	19
9	澳大利亚	79	韩国	19
10	意大利	68	法国	18

为密切。

2.4 中美 CRISPR/Cas 论文作者所属机构情况

通过对论文作者所属机构进行分析,可以了解某一领域不同机构的研究水平^[12]。如图3所示,全球CRISPR/Cas论文数量最多的前15位研究机

构全部为大学和科研院所,说明大学和科研院所是CRISPR/Cas领域基础研究的主要力量。在全球前15位机构中,美国机构占了12所,中国只占了2所。

美国发表CRISPR/Cas论文数量最多的前15位机构中,大学占了8所,科研院所占了7所,其中

实力最强的是美国加州大学系统和美国哈佛大学, 发表 CRISPR/Cas 论文数量分别是 574 篇和 531 篇。美国其他机构发表的 CRISPR/Cas 论文数量均在百篇以上, 与前两所机构相比差距不大, 表明美国 CRISPR/Cas 领域的前 15 位机构研究力量较为均衡。中国 CRISPR/Cas 领域的前 15 位机构中仅有 4 所是科研院所, 其他全部为大学, 其中实力最强的是中国科学院, 发表 CRISPR/Cas 论文数量多达 405 篇。除了中国科学院大学和中国科学院上海生命科学研究院以外, 中国其他机构发表的 CRISPR/Cas 论文数量均在百篇以下, 与中国科学院相比差距较大,

表明中国在 CRISPR/Cas 领域的前 15 位机构研究力量比较集中。在全球 CRISPR/Cas 领域前 15 位机构中, 中国只有中国科学院和中国科学院大学进入, 全球排名分别为第 3 和第 12, 表明中国该领域的研究机构实力相对美国较弱。此外, 中国科学院在 CRISPR/Cas 领域开展的中美合作最为密切, 近 10 年共合作发表 CRISPR/Cas 论文 108 篇。美国加州大学系统和美国哈佛大学在该领域开展的中美合作也比较多。值得注意的是, 中国科学院大学发表的 CRISPR/Cas 论文数量虽然较多, 但中美合作的论文数量相对较少。

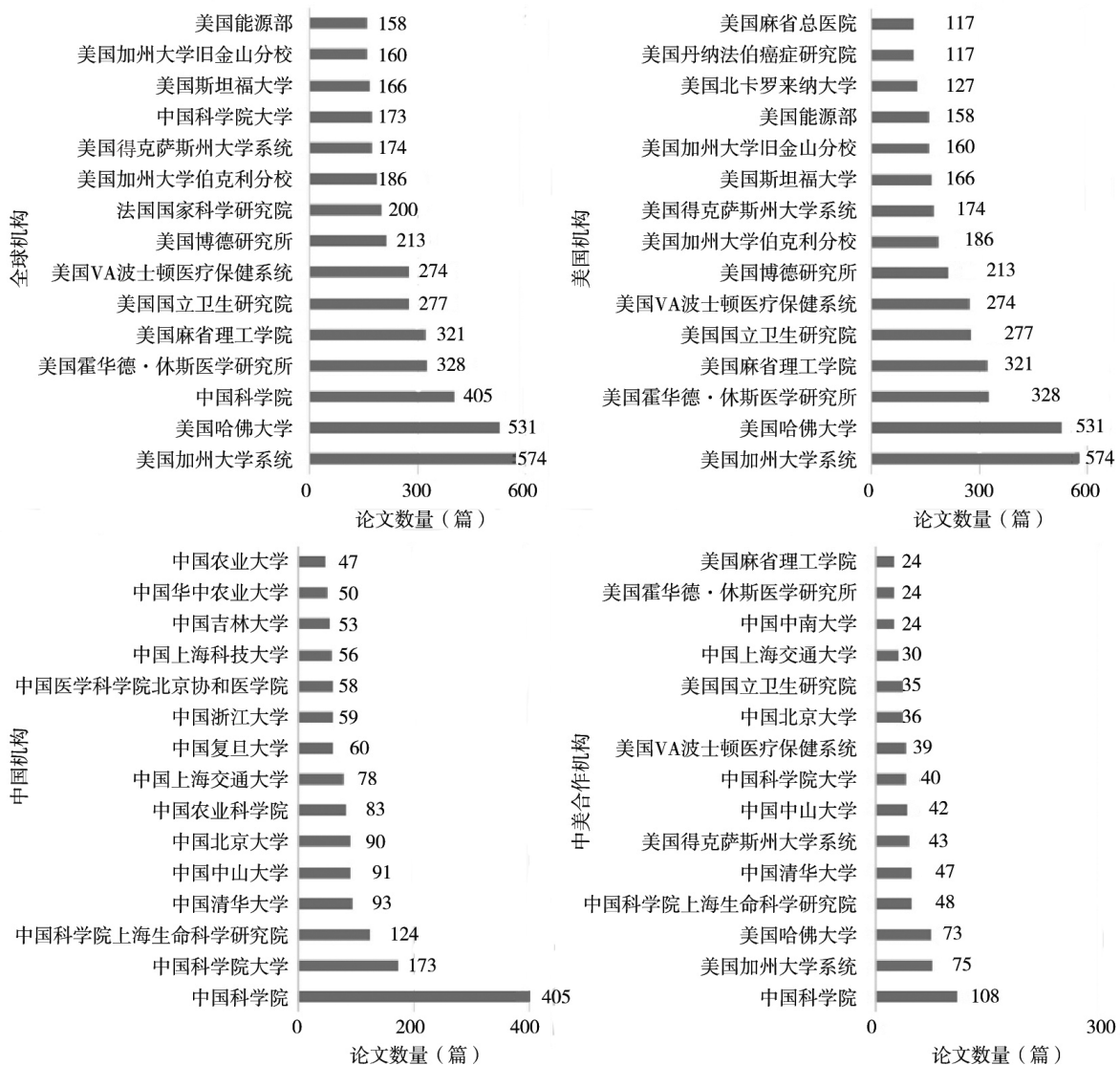


图 3 CRISPR/Cas 技术领域论文的前 15 位机构排名情况

2.5 中美 CRISPR/Cas 高被引论文情况

通过统计一个国家在某一领域的高被引论文数量,能够了解这个国家在该领域的科研成果质量和学术影响力。近些年来,论文产出量及影响力,尤其是高水平论文产出状况,正成为衡量一个国家科研能力最基本和最常用的指标。根据二八定律^[13],将 CRISPR/Cas 技术领域的论文中被引次数占前 20% 的论文设定为该领域的高被引论文,如果前 20% 之后有论文的被引次数与第 20% 篇论文相等,则同样收录。经计算得到,美国和中国在 CRISPR/Cas 领域的高被引论文数量分别为 764 篇和 363 篇。

从表 2 可见,美国和中国 CRISPR/Cas 论文占全球相关论文的比例分别为 49.42% 和 23.53%,两国

高被引论文占有高被引论文的比例分别为 49.45% 和 23.50%,可见中美两国 CRISPR/Cas 论文占全球相关论文的比例与其高被引论文占有高被引论文的比例基本相同,表明中美两国在 CRISPR/Cas 基础研究领域的影响力很高。H 指数是学术界最常用的科研成就评价指标,美国 CRISPR/Cas 论文的 H 指数高达 156,是中国的 2 倍,说明美国 CRISPR/Cas 论文的学术影响力更高。此外,中美合作的 CRISPR/Cas 论文的篇均被引次数比中国独著论文的篇均被引次数多,高被引论文的篇均被引次数也比中国高被引论文的篇均被引次数多,表明中美合作的研究成果质量较高,能产生较大的学术影响力。

2.6 中美 CRISPR/Cas 顶尖研究机构情况

以表征研究质量和学术影响力的高被引论文数

表 2 SCI 收录的 CRISPR/Cas 技术领域的高被引论文情况

	论文数量 (篇)	论文占全 球比例 (%)	总被引次 数(次)	篇均被引 次数(次)	H 指数	高被引论 文数量 (篇)	高被引论 文占有 高被引论 文比例 (%)	高被引论 文占该国 所发表论 文比例 (%)	高被引论 文的被引 次数(次)	高被引论 文的篇均 被引次数 (次)
美国	3 820	49.42	134 006	35.08	156	764	49.45	20	113 269	148.26
中国	1 819	23.53	31 938	17.56	77	363	23.50	19.96	27 621	76.09
中美合作	582	7.53	17 246	29.63	56	116	7.51	19.93	15 077	129.97

量为横坐标,以表征科研成果平均水平的篇均被引次数为纵坐标,将中美两国 CRISPR/Cas 领域的顶尖研究机构做成散点图并划分为 4 个区域^[14]。如图 4 所示,将高被引论文数量及篇均被引次数都高者列为第一区,如哈佛大学;将仅高被引论文数量多者列为第二区,如加州大学系统和中国科学院;将仅篇均被引次数高者列为第三区,如博德研究所;将高被引论文数量及篇均被引次数均不高者列为第四区。

从图 4 可以看到,美国在 CRISPR/Cas 领域的顶尖机构在 4 个区都有分布。第一区全部为美国机构,分别是哈佛大学、麻省理工学院和霍华德·休斯医学研究所,这 3 家顶尖机构的高被引论文数量和篇均被引次数都很高。中国 CRISPR/Cas 领域的顶尖机构主要集中在第四区,个别机构在第二区和第三区,说明中国顶尖机构在 CRISPR/Cas 领域的科研实力相对美国较弱。位于第二区的中国顶尖机构是中国科学院,其高被引论文数量最多,

多达 81 篇,但篇均被引次数不高,仅 22.71 次/篇;位于第三区的中国顶尖机构是清华大学,其篇均被引次数最高,高达 83.03 次/篇,但高被引论文数量较少,仅 18 篇。

3 结语

3.1 中美 CRISPR/Cas 基础研究发展迅速,美国发文量处于世界领先地位

从表征研究水平和发展趋势的论文逐年发表数量来看,全球 CRISPR/Cas 论文数量近 10 年呈逐年上升趋势,发展速度不断加快,CRISPR/Cas 基础研究开始步入全面发展的重要时期。从 2012 年开始,全球 CRISPR/Cas 论文数量增长幅度明显加大。究其原因,可能是因为 M. Jinek 等^[15]在 2012 年率先证实了 Cas9 可以在人工合成的 sgRNA 引导下对靶标 DNA 序列进行特异切割,CRISPR 技术因此成为了当时基因编辑技术的研究主流。中美发展趋势与

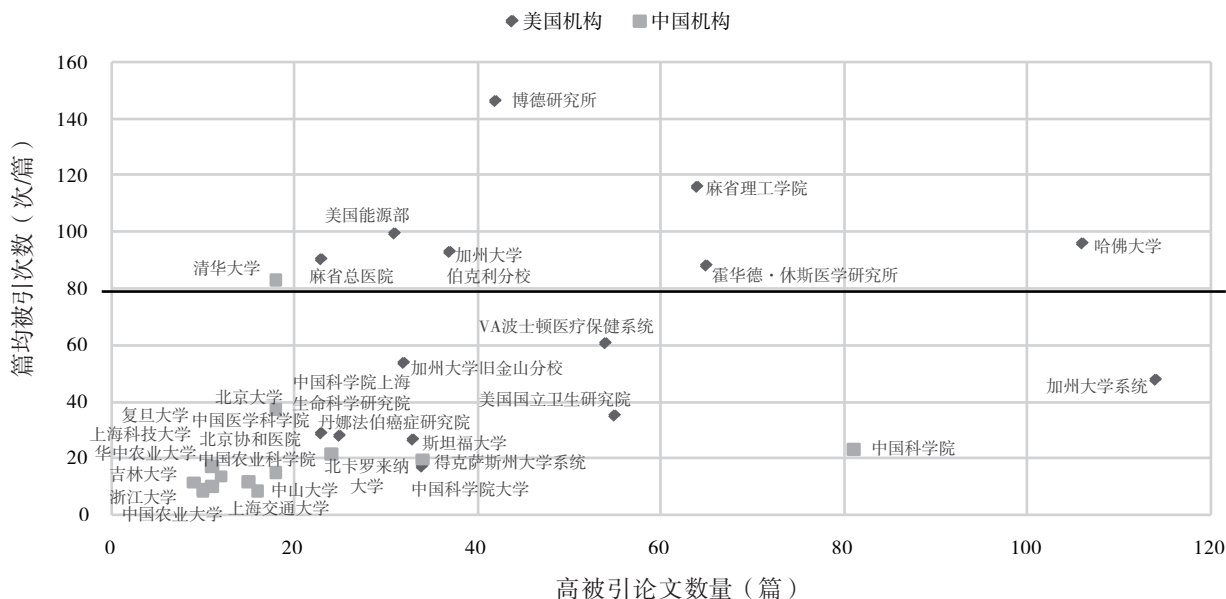


图 4 CRISPR/Cas 技术领域中美顶尖机构情况

全球趋势高度一致, 表明中美紧随全球基因编辑技术的发展主流, 积极开展对 CRISPR 领域的研究。从表征全球研究布局的世界国家排名来看, 中美两国是 CRISPR/Cas 领域基础研究的中坚力量。美国 CRISPR/Cas 发文量高达 3 820 篇, 处于世界领先地位。中国 CRISPR/Cas 发文量为 1 819 篇, 位居世界第二。中美两国 CRISPR/Cas 发文量远高于世界其他国家。

3.2 美国 CRISPR/Cas 机构实力强劲且分布均衡, 中国研究力量较为集中

从表征某一领域不同机构研究水平的作者所属机构来看, 全球 CRISPR/Cas 领域前 15 位研究机构中, 美国机构占了 12 所。美国发表 CRISPR/Cas 论文数量最多的机构是加州大学和哈佛大学, 在全球机构中处于领军地位。单从数量上看, 美国机构的实力不容小觑。美国其他机构发表的 CRISPR/Cas 论文数量与前两所机构相比差距不大, 表明美国该领域的前 15 位机构研究力量分布均衡。中国其他机构发表的 CRISPR/Cas 论文数量与中国科学院相比差距较大, 表明中国该领域的基础研究以中国科学院为主要力量。从表征研究质量的高被引论文数量和表征科研成果平均水平的篇均被引次数综合来看, 美国顶尖机构的研究实力非常强劲, 第一、二、三区机构基本以美国机构为主。值得注意的是, 美国加州大学系统的 CRISPR/Cas 论文数

量虽然排名第一, 但是其篇均被引次数并不高。中国顶尖机构主要集中在第四区, 只有中国科学院和清华大学两所机构位于第二、第三区。可见, 中国在 CRISPR/Cas 领域的研究力量比较集中。

3.3 中美两国 CRISPR/Cas 领域科研合作密切, 中美合作成果认可度高

从表征某一领域不同国家和机构合作密切程度的合作分布情况来看, 中美两国在 CRISPR/Cas 领域的科研合作最密切。中美两国近 10 年共合作发表 CRISPR/Cas 论文 582 篇, 但中国与其他国家合作发表的 CRISPR/Cas 论文数量均不足 100 篇, 可见中国与美国之间的合作非常紧密。中美合作成果数量最多的机构是中国科学院, 共合作发表相关论文 108 篇。此外, 美国加州大学系统和哈佛大学在 CRISPR/Cas 领域开展的中美合作也比较多。从表征科研成果平均水平的篇均被引次数来看, 中美合作的 CRISPR/Cas 论文篇均被引次数比中国独著论文的多, 说明中美合作的 CRISPR/Cas 论文认可度较高。

3.4 中美 CRISPR/Cas 科研成果质量较高, 美国论文学术影响力更大

从表征研究质量和学术影响力的论文引用情况来看, 中美 CRISPR/Cas 论文占全球相关论文的比例与其高被引论文占有所有高被引论文的比例基本相同, 表明中美两国在 CRISPR/Cas 领域的科研成果质量较

高。但美国的 CRISPR/Cas 论文总被引次数和高被引论文的被引次数远高于中国,说明美国在 CRISPR/Cas 领域的研究认可度更高。此外,美国 CRISPR/Cas 论文的 H 指数高达 156,而中国的 H 指数只有 77,可见美国 CRISPR/Cas 论文的学术影响力更大。■

参考文献:

- [1] 李想, 崔文涛, 李奎. 基因编辑技术及其应用的研究进展 [J]. 中国畜牧兽医, 2017, 44 (8) : 2 241-2 247.
- [2] Petersen B, Niemann H. Advances in genetic modification of farm animals using zinc-finger nucleases (ZFN)[J]. Chromosome Res, 2015, 23(1): 7-15.
- [3] Karakikes I, Termglinchan V, Cepeda DA, et al. A comprehensive TALEN-based knockout library for generating human-induced pluripotent stem cell-based models for cardiovascular diseases[J]. Circ Res, 2017, 120(10): 1 561-1 571.
- [4] Javed MR, Sadaf M, Ahmed T, et al. CRISPR-Cas system: History and prospects as a genome editing tool in microorganisms[J]. Curr Microbiol, 2018, 75(12): 1 675-1 683.
- [5] 陈怡李, 姚书忠. CRISPR/Cas9 基因编辑技术的应用研究进展 [J]. 国际生殖健康 / 计划生育杂志, 2017, 36 (6) : 482-487.
- [6] 杨梦婷, 高嘉琪, 倪菊华. CRISPR-Cas9 技术的发展与临床应用前景 [J]. 生理科学进展, 2017, 48 (6) : 463-468.
- [7] 孙轶楠, 杜建, 唐小利. 基于科学论文和专利信息的基因测序技术领域创新态势分析 [J]. 中华医学图书情报杂志, 2017, 26 (5) : 31-37.
- [8] 孟洋. 基于 Web of Science 数据库高水平论文的中日大学科研实力比较与探析 [J]. 科教导刊, 2018 (17) : 11-12, 81.
- [9] 曹学伟, 高晓巍, 陈锐. 基于文献计量分析的基因编辑技术发展研究 [J]. 全球科技经济瞭望, 2018, 33 (4) : 60-72.
- [10] 赵润州, 徐卸古. 基于文献计量的 CRISPR-Cas 基因组编辑技术可视化分析 [J]. 医学研究生学报, 2016, 29 (6) : 639-644.
- [11] 王新明, 丁敬达. 科研论文的著者合作模式研究综述 [J]. 现代情报, 2018, 38 (8) : 172-177.
- [12] 盛宇平, 王树贵, 盛春蕾. 作者机构指标纳入学术期刊评价重要指标的探讨 [J]. 现代情报, 2018, 38 (5) : 117-119, 139.
- [13] 吴凤芝, 席宁, 尤海燕, 等. 基于“二八定律”分析《现代中医临床》高被引论文 [J]. 现代中医临床, 2017, 24 (5) : 52-56.
- [14] 周群, 韩涛, 左文革, 等. 基于学科前沿性视角的科研机构评测研究与实证 [J]. 现代情报, 2018, 38 (4) : 65-70, 76.
- [15] Jinek M, Chylinski K, Fonfara I, et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity[J]. Science, 2012(337): 816-821.

A Comparative Study of Chinese and American CRISPR/Cas Fields Based on Paper Analysis

SUN Jing-song, ZHENG Yan-ning, YUAN Fang

(Institute of Scientific and Technical Information of China, Beijing 100038)

Abstract: Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-associated proteins (CRISPR/Cas) is a research hotspot in the field of gene editing and it is widely used in life science research, disease model construction, drug research and development, and agricultural production. Based on the data of CRISPR/Cas scientific papers from 2008 to 2018, this paper uses the method of bibliometrics to analyze the Sino-US CRISPR/Cas technology from the perspectives of the number of publications, major country rankings, distribution of cooperative countries, major research institutions and highly cited papers. The basic research progress in the field is comparatively analyzed, aiming to provide certain reference for the government, scientific research institutions and related enterprises.

Key words: United States; China; paper analysis; CRISPR/Cas technology