

利用GoPubMed对CRISPR相关文献的计量学分析

边富宁 李文革 李先平 卢金星

【导读】 为了解CRISPR (clustered regularly interspaced short palindromic repeats) 相关研究现状, 笔者采用GoPubMed数据库的在线分析软件, 对CRISPR相关文献进行计量分析。共检索到文献758篇。CRISPR相关研究文献呈逐年上升趋势, 其中美国在CRISPR的文献数量和影响力均居首位, 中国以北京地区发表文献最多(30篇)。PLOS One 刊登相关文献47篇, 占发表总量的6.20%, 居首位。CRISPR研究主题词涉及基因、基因组学、免疫等。高产作者为Horvath, 共发表CRISPR相关文章11篇。利用GoPubMed可较好反映CRISPR研究的现状和发展趋势。中国CRISPR研究文献在整体质量和数量上均有待进一步提高。

【关键词】 文献计量分析; GoPubMed; CRISPR

Application of GoPubMed in bibliometric analysis of literature on CRISPR *Bian Funing, Li Wenge, Li Xianping, Lu Jinxing. State Key Laboratory for Communicable Diseases Prevention and Control, Chinese Center for Disease Control and Prevention, Beijing 102206, China*

Corresponding authors: Li Xianping, Email: lixianping@icdc.cn; Lu Jinxing, Email: lujinxing@icdc.cn
This work was supported by a grant from the National Key Technology Support Program (No. 2012BAI11B05).

【Key words】 Bibliometric analysis; GoPubMed; Clustered regularly interspaced short palindromic repeats

GoPubMed是以PubMed为数据源, 以语义检索技术为支撑, 对检索结果进行分类分析的数据挖掘工具, 可有效精简检索结果, 高效检索目标文献, 对于文献统计及研究热点的追踪提供了很大帮助^[1]。CRISPR (clustered regularly interspaced short palindromic repeats) 规律间隔成簇短回文重复序列, 是细菌和古细菌在对抗噬菌体的生存斗争中进化出的获得性免疫系统, 能对噬菌体或者外源性DNA进行切割和降解^[2]。CRISPR结构最初(1987年)由日本科学家Ishino于大肠埃希菌中发现, 当时并未引起关注, 2002年由Jansen等将其命名为CRISPR^[3,4]。2007年Barrangou等^[2]通过实验证实CRISPR系统能使宿主抵抗噬菌体、质粒等外来的DNA入侵的免疫能力。2013年CRISPR/Cas9作为基因编辑工具首次被Science报道后, 许多科学家投入到这个充满魅力的技术上^[5]。2013年在世界三大权威期刊(*Science*、*Nature*和*Cell*)共刊出CRISPR相关文章14篇。2013年12月Science公布了其评出的“2013年度十大科学突破”, CRISPR技术位列其中^[6]。为了解CRISPR研究的文献增长规律和发展趋势, 为从事CRISPR相关研究人员提供参考, 笔者采用GoPubMed对CRISPR相关文献进行分析。

1. 数据来源: 登录GoPubMed网站(www.GoPubMed.com), 输入检索词“Clustered Regularly Interspaced Short Palindromic Repeats”[mesh]进行检索, 检索时间为2014年5月27日。通过GoPubMed网站统计功能, 分析被PubMed收录的所有CRISPR相关文献。

2. 结果与分析: 经检索, PubMed收录CRISPR相关文章共758篇, 检索结果均纳入统计分析。

(1) 文献年度分布: 自2002年第一篇CRISPR相关论文发表后, 2002—2012年文献数量经过缓慢增长, 于2012年始呈快速增长, 2012年全年PubMed收录CRISPR相关文章123篇, 2013年收录文献较2012年成倍增长, 共收录文章241篇, 2014年至笔者检索时已收录文献193篇(图1A)。据GoPubMed统计, CRISPR相对研究兴趣指数(relative research interest, RRI)^[7]逐年提高, 从2012年起增长幅度较大, 2012年RRI约为0.000 240, 2013年预期可达0.000 380, 反映出CRISPR在所有研究领域中的关注度逐年快速增长。RRI(按当年PubMed中该研究领域加权论文数量在所有研究领域加权论文集合中所占比例计算出RRI)可较好反映研究活跃程度, 比文献的绝对数量更重要(图1B)。

(2) 国家和城市文献分布: CRISPR相关研究主要分布于北美洲、欧洲和亚洲(图2A)。按国家排名, 美国发表CRISPR相关文章288篇(27.99%), 位居首位(图2B)。中国和德国分别发表59和57篇, 排名第二、三位。如按城市分类统计文献数量, 北京(中国)发表30篇排第一位, 伯克利(美国)和哥本哈根(丹麦)分列第二位和第三位(图2C)。

DOI: 10.3760/cma.j.issn.0254-6450.2014.12.020

基金项目: 国家重点科技支撑计划(2012BAI11B05)

作者单位: 102206 北京, 中国疾病预防控制中心传染病预防控制所
传染病预防控制国家重点实验室

通信作者: 李先平, Email: lixianping@icdc.cn; 卢金星, Email: lujinxing@icdc.cn

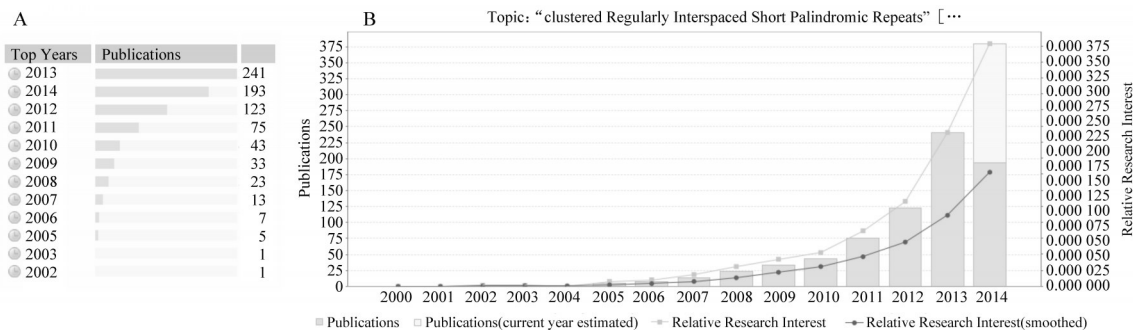


图1 CRISPR 相关文献发表的年度分布

(3)来源期刊分布:刊载CRISPR研究论文的主要来源期刊见表1。其中 *PLoS One* 为刊登CRISPR研究论文最多的杂志。*Science*、*Nature* 等国际权威杂志也刊登CRISPR相关研究文献,其中 *Science* 刊登15篇,*Nature* 刊登12篇,*Cell* 刊登10篇。

(4)研究主题:目前CRISPR研究的主题集中在CRISPR结构本身的研究以及相关基因组学研究(表2)。表明这些研究主题可代表该领域的研究趋势和热点。CRISPR作为细菌体的免疫系统,大量发表文献可出现“immunity/immunization”和“CRISPR-Cas systems”等主题词。随着CRISPR/Cas9作为基因编辑工具首次被 *Nature* 报道后,预期将会出现基因编辑、基因敲除等新的主题词^[5]。

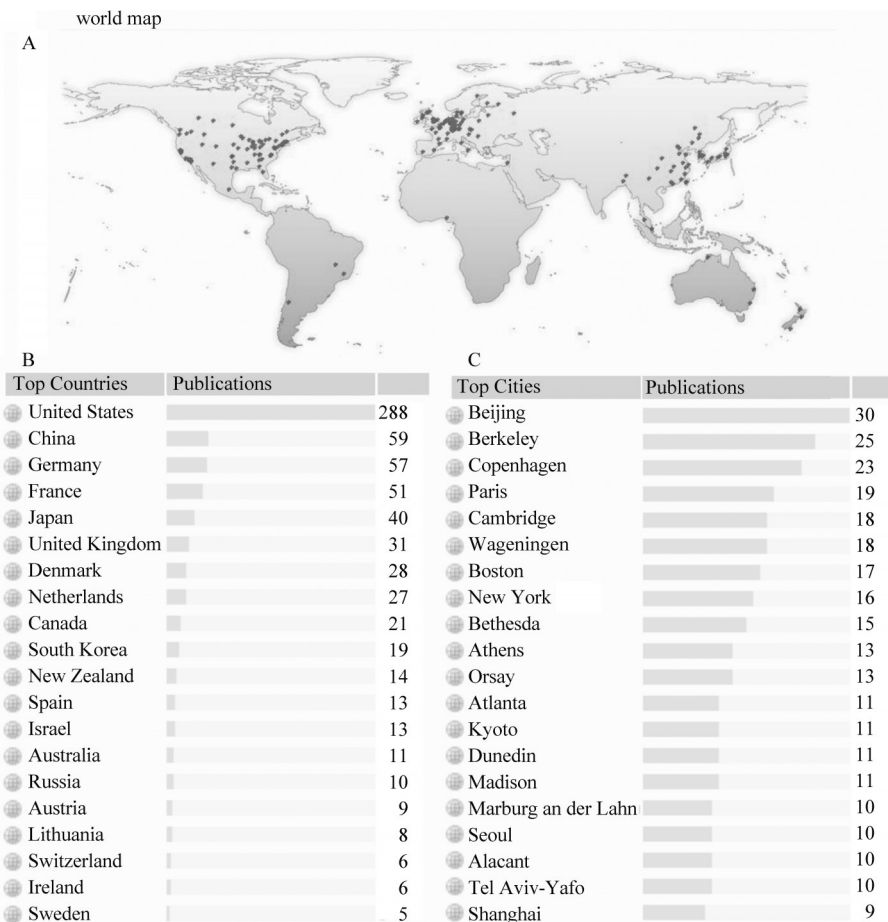
(5)高产作者及相互关系:分析表明 Horvath 发表CRISPR相关文章11篇,排名第一;Vergnaud 发表8篇排名第二(图3A)。CRISPR文献的作者主要形成两大合作网络,最大的合作网络是以 Horvath 为核心,据图3B中线条的复杂性和粗细,可知研究者间合作的密切程度。

(6)中美两国发表文献期刊比较:美国和中国发表CRISPR相关文章数量为前2名,发表论文的质量可由刊登杂志的影响因子反映出。美国在 *PLoS One*、*Proc Natl Acad Sci USA* 和 *J Bacteriol* 发表相关文章数量均为12篇,三者并列第一;在 *Science* 发表10篇,*Cell* 发表8篇,*Nature* 发表6篇(图4A)。中国发表的文献数量分布前3名分别是 *Cell Res*、*PLoS One* 和《微生物学报》。其中 *Cell Res* 和《微生物学报》为中国出版期刊,*Cell Res* 被SCI收录(2012年IF为10.526),刊登6篇论文;《微生物学报》未被SCI收录,刊登4篇。在 *Nature* 仅刊

登1篇。

3. 评价:GoPubMed 是优秀的智能检索工具。不仅可以针对某研究课题进行分析,总结出该课题领域的最新相关情况,还可以对一个研究机构发表的文献进行分析,从而揭示其研究方向、核心作者、研究趋势等^[1]。

CRISPR 修饰系统为最新的基因编辑工具,较之传统的基因修饰技术,如锌指核酸酶(zinc-finger nucleases)、转录激活因子样效应物核酸酶(transcription activator-like effector nucleases),具有操作简单、成本较低和多基因同时操作等优势,由于其适用范围广泛,目前已应用于动物模型、植物和基



注: A. 世界分布; B. 国家分布及其排名; C. 城市分布及其排名

图2 CRISPR 文献数量的地域分布

表 1 CRISPR 研究论文主要来源期刊及其比例

排名	来源期刊	文献篇数	比例 (%)
1	<i>PLoS One</i>	47	6.20
2	<i>Nucleic Acids Res</i>	36	4.75
2	<i>Rna Biol</i>	36	4.75
3	<i>J Bacteriol</i>	26	3.43
4	<i>Mol Microbiol</i>	22	2.90
5	<i>Biochem Soc Trans</i>	20	2.64
6	<i>Proc Natl Acad Sci USA</i>	17	2.24
6	<i>Appl Environ Microbiol</i>	17	2.24
7	<i>Bmc Genomics</i>	16	2.11
8	<i>Science</i>	15	1.98
9	<i>J Biol Chem</i>	13	1.72
10	<i>Nature</i>	12	1.58
11	<i>Mol Cell</i>	11	1.45
11	<i>PLoS Genet</i>	11	1.45
11	<i>Environ Microbiol</i>	11	1.45
12	<i>Nat Methods</i>	10	1.32
12	<i>Nat Biotechnol</i>	10	1.32
12	<i>Cell</i>	10	1.32

表 2 CRISPR 研究主题中前 20 个主题词及文献分布

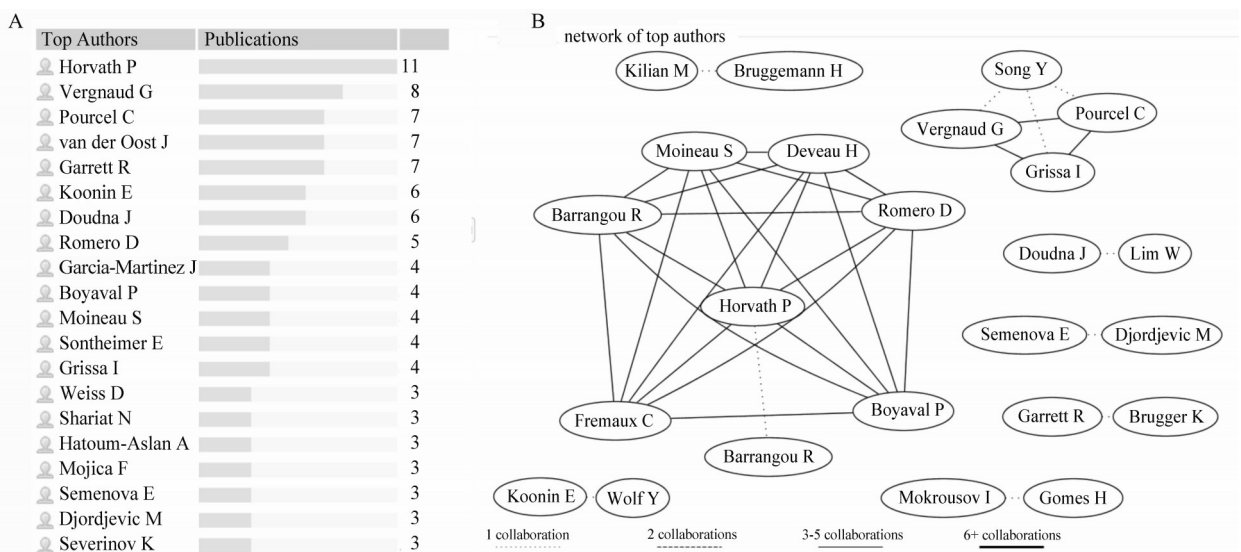
高频主题词	文献篇数
clustered regularly interspaced short palindromic repeats	758
cluster analysis	482
genome	445
genomics	442
genes	426
maintenance of CRISPR repeat elements	424
DNA	327
immunity	289
immunization	288
proteins	287
bacteria	252
RNA	234
CRISPR-Cas systems	230
base sequence	225
catalase	223
breast cancer anti-estrogen resistance protein 1	222
cycloartenol synthase	222
probable O-acetyltransferase CAS1	222
CAS1 domain-containing protein 1	222
transcription factor castor	222

础医学等多个领域^[8-12]。

分析 CRISPR 相关研究文献发表的时间提示呈逐年上升趋势,且增幅较大,自 2012 年起开始大幅度增长。反映了当前对 CRISPR 的研究处于起步阶段,其论文产出不断增加,同时 CRISPR 正在由免疫机制研究到基因编辑技术的应用^[9,11],作为当前最有效的基因编辑技术其文献产出必然会有更大幅度的增长。分析论文分布区域,主要集中在美国、欧洲和中国,反映出我国对 CRISPR 研究的关注程度也在提高。北京以 30 篇论文名列城市排名榜首,其中不乏高影响力的国际期刊,说明北京地区在生命科学技术的研究中具有一定影

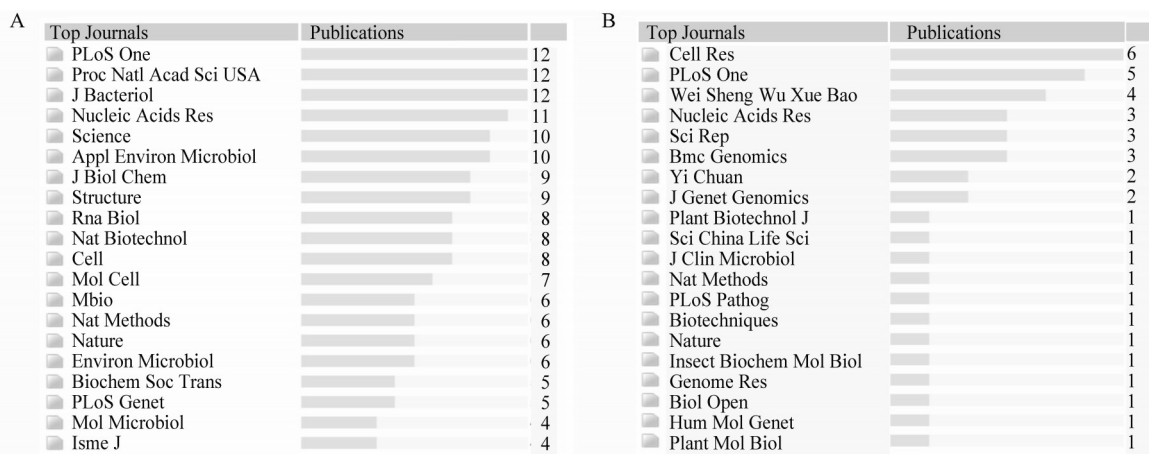
响力。从文献分布可见,Science、Nature 和 Cell 国际三大顶尖科技期刊共发表 CRISPR 相关研究 37 篇,反映出对 CRISPR 较高的关注度。

我国发表论文总数量排名第二,但分析中美两国发表论文的影响力,在 Science、Nature 和 Cell 上共刊登 CRISPR 相关研究 37 篇,其中美国发表 24 篇,中国仅有 1 篇。说明在 CRISPR 相关研究中我国不仅在论文数量上与美国有差距,且在论文影响力方面的差距更为明显。提示我国在研究上应注重创新性,提高论文水平。



注: A. 前 20 名高产作者; B. 高产作者研究关系网络

图 3 CRISPR 相关文献中高产作者及其研究关系分布



注: A 美国; B 中国

图 4 中美两国发表 CRISPR 相关文献期刊

参 考 文 献

[1] Doms A, Schroeder M. GoPubMed: exploring PubMed with the Gene Ontology [J]. Nucleic Acids Res, 2005, 33 (Web Server Issue): W783-786.

[2] Barrangou R, Fremaux C, Deveau H, et al. CRISPR provides acquired resistance against viruses in prokaryotes [J]. Science, 2007, 315(5819): 1709-1712.

[3] Ishino Y, Shinagawa H, Makino K, et al. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product [J]. J Bacteriol, 1987, 169(12): 5429-5433.

[4] Jansen R, Embden JD, Gastra W, et al. Identification of genes that are associated with DNA repeats in prokaryotes [J]. Mol Microbiol, 2002, 43(6): 1565-1575.

[5] Cong L, Ran FA, Cox D, et al. Multiplex genome engineering using CRISPR/Cas systems [J]. Science, 2013, 339 (6121): 819-823.

[6] Coontz R. Science's Top 10 Breakthroughs of 2013 [EB/OL]. [2013-12-19]. <http://news.sciencemag.org/2013/12/sciences-top-10-breakthroughs-2013>.

[7] Bailey J. An assessment of the use of chimpanzees in hepatitis C research past, present and future: 1. Validity of the chimpanzee model [J]. Altern Lab Anim, 2010, 38(5): 387-418.

[8] Hsu PD, Lander ES, Zhang F. Development and applications of CRISPR-Cas9 for genome engineering [J]. Cell, 2014, 157(6): 1262-1278.

[9] Feng Z, Mao Y, Xu N, et al. Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in arabidopsis [J]. Proc Natl Acad Sci USA, 2014, 111(12): 4632-4637.

[10] Wu Y, Liang D, Wang Y, et al. Correction of a genetic disease in mouse via use of CRISPR-Cas9 [J]. Cell Stem Cell, 2013, 13(6): 659-662.

[11] Bondy-Denomy J, Pawluk A, Maxwell KL, et al. Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system [J]. Nature, 2013, 493(7432): 429-432.

[12] Xu R, Li H, Qin R, et al. Gene targeting using the Agrobacterium tumefaciens-mediated CRISPR-Cas system in rice [J/OL]. Rice (NY), 2014, 7(1): 5. <http://www.thricejournal.com/content/7/1/5>. DOI: 10.1186/s12284-014-0005-6.

(收稿日期: 2014-06-20)
(本文编辑: 张林东)

• 征 订 启 事 •

本刊 2015 年征订启事

《中华流行病学杂志》是由中华医学会主办的流行病学及其相关学科的高级专业学术期刊、国内预防医学和基础医学核心期刊、国家科技部中国科技论文统计源期刊, 2004—2012 年被中国科学技术信息研究所定为“百种中国杰出学术期刊”, 并被美国国立图书馆医学文献联机数据库 (Medline) 和美国化学文摘社 (CAS) 收录。读者对象为医学 (预防医学、临床医学、基础医学及流行病学科研与教学) 和健康相关学科的科研、疾病控制、临床、管理和教学工作者。刊稿范畴: 重点或新发传染病现场调查与控制; 慢性病的病因学及流行病学调查 (含社区人群调查)、干预与评价; 伤害的流行病学与防控; 环境污染与健康; 食品安全与食源性疾病; 临床流行病学和循证医学; 流动人口与疾病; 行为心理障碍与疾病; 分子和遗传流行病学与疾病控制; 我国西部地区重点疾病的调查与控制; 理论流行病学; 流行病学教学与实践等。本刊设有述评, 论著 (原著) 包括现场调查、监测、实验室研究、临床研究、基础理论与方法、疾病控制、国家课题总结、国外杂志华人研究导读 (科海拾贝)、文献综述、问题与探讨等重点栏目。

全年出版 12 期, 每期定价 20 元 (含邮费), 全年 240 元, 由全国各地邮局统一订阅, 邮发代号: 2-73。本刊编辑部常年办理邮购。地址: 北京市昌平区昌百路 155 号传染病所 B115《中华流行病学杂志》编辑部, 邮编: 102206, 电话 (传真): 010-58900730, Email: zhlx1981@sina.com 欢迎广大读者踊跃投稿 (<http://chinaepi.icdc.cn>), 积极订阅。

本刊编辑部