

基于文献计量分析的基因编辑技术发展研究

曹学伟, 高晓巍, 陈 锐

(中国科协创新战略研究院, 北京 100012)

摘 要:“基因编辑”技术指能够让人类对目标基因进行“编辑”, 实现对特定 DNA 片段的敲除、加入等的新兴生物学技术。为了解中国该领域的发展态势, 本文运用文献计量法, 基于 2007—2017 年 Scopus 基因编辑技术领域论文数据, 对该技术领域的国际研究进展进行了整体分析, 与国际主要国家对比分析了中国该领域 SCI 论文数量及其被引用情况, 对核心作者、核心期刊、主要国家、专业研究机构及主要研发团队、申请专利数量和经济影响、主要分布学科以及研究前沿热点进行系统分析。旨在对我国和世界基因编辑技术领域的研究现状进行整体把握和分析, 以期更好地应对新兴产业创新发展带来的机遇和挑战。

关键词: 基因编辑; 文献计量; 核心作者; 研究机构; 专利分析; 研究前沿和热点

中图分类号: R392 **文献标识码:** A **DOI:** 10.3772/j.issn.1009-8623.2018.04.010

20 世纪 70 年代 DNA 重组技术的发展标志着生物学新时代的开始。随后, 基因组工程技术 (Genome Engineering Technologies) 的一系列最新进展引发了生物研究的新革命。基于基因同源重组 (Homologous Recombination, HR) 方法开发的基因靶向技术 (Gene Targeting) 可以用于删除某一基因、去除外显子或导入点突变, 通过改变生物体某一内源基因而实现对此基因的功能研究。然而, 尽管 HR 介导的基因靶向产生的定向精确度很高, 但是目标重组事件发生概率极低。为了克服这些挑战, 近年来研究人员开发了一系列基因组编辑技术, 能够有针对性、有效地对真核生物特别是哺乳动物进行基因修饰。

“基因编辑”技术指能够让人类对目标基因进行“编辑”, 实现对特定 DNA 片段的敲除、加入等的新兴生物学技术。作为目前能够直接对 DNA 进行编辑的基因组工程技术中最具应用前景的技术手段, 基因编辑技术在医学 (基因靶向治疗、

药物研发)、生物学 (动物模型、遗传变异) 以及生物技术 (能源、食品、材料) 等方面展示出了巨大的潜力^[1]。基因编辑技术主要以锌指核酸酶 (Zinc-Finger Nucleases, ZFN)^[2]、转录激活因子样效应因子核酸酶 (Transcription Activator-Like Effector Nucleases, TALEN)^[3] 和通过 sgRNA 介导的基于成簇的规律间隔的短回文重复序列 Cas 蛋白的 DNA 核酸内切酶 (Clustered Regularly Interspaced Short Palindromic Repeats, CRISPR/Cas9)^[4] 等为代表, 相关研究人员在国内外顶级期刊上发表了大批研究成果, 在基础研究、基因治疗和遗传改良等方面取得了多项突破性进展。2014 和 2016 年美国麻省理工学院科学研究评论分别将基因编辑技术和基因编辑技术在植物上的应用作为年度科技前沿十大热点之一。近年来基因编辑技术已经成为生命科学领域重要的颠覆性技术成果之一。

文献计量学方法是数据挖掘分析的重要方法^[5]。近几十年来文献计量学方法由于其客观性和定量研

第一作者简介: 曹学伟 (1987—), 女, 博士, 主要研究方向为科技管理文献计量。

收稿日期: 2018-01-14

研究的特性而得到越来越广泛的应用。利用定量分析工具对各学科研究文献数量进行科学计量分析和比较,既可以对国家、高校/科研机构等进行学科分析评估,动态分析国内外各研究机构不同学科的发展趋势以及学科差距;又能够了解学科领域内核心作者的关键信息;同时还能对学科的发展趋势进行预测,既能够为政策制定者遴选和制定科技战略规划提供重要指标和参考,又能够使研究人员了解和洞察当前科学技术领域^[6]。

为了解近 10 年来基因编辑领域快速发展的趋势及未来发展方向,对标国内当前发展现状,本文利用文献计量学原理对相关文献进行分析,研究文献的分布结构数量关系变化规律;对核心作者、核心期刊、主要国家和地区、专业研究机构及主要研发团队、申请专利数量和经济影响、主要分布学科等要素进行统计分析;同时还利用 VOSviewer 软件构建关键词词频共现地图,对基因编辑领域的基本概况、发展动态进行整体把握和分析,进一步了解其研究热点。

1 数据来源及研究方法

1.1 数据来源

本文使用爱思唯尔的 Scopus 核心数据库,对收录其中的有关基因编辑的文献进行主题检索,检索时间为 2007—2017 年(截止时间为 2017 年 9 月 12 日)。检索式为(“gene editor” or “ZFN” or “TALEN” or “CRISPR”),共检索到文献 8 251 篇,为了更好地突出文献蕴含的信息,对 8 251 篇文献进行筛选,取出其中类型为 Article 的文献 5 804 篇,Review 1 094 篇,Article in Press 172 篇,Conference Paper 103 篇,共计 7 173 篇,对这 7 173 篇文献进行文献计量学分析。

1.2 分析方法

本研究采用 Scopus 数据,并利用基于 Scopus 数据库的 SciVal 分析平台,根据概览模块(Overview Module)、标杆模块(Benchmarking Module)、协作模块(Collaboration Module)以及趋势模块(Trends Module)中提供的相关指标数据,重点分析研究基因编辑领域的研究产出力、引文影响力、合作状况和学科制度等。

VOSviewer^[7]是一款用来构建和查看文献计量

图谱的分析软件,基于文献的共引和共被引原理,可用于绘制各个知识领域的科学图谱。本研究主要利用其中的 text corpus 模块分析 7 173 篇文献中的标题和摘要中的词汇共现,基于同现词网络构建关键词术语地图、热点研究地图等。

2 结果与分析

2.1 文献量分析

从期刊文献发表的数量可以最直观地看出某一领域目前的研究水平、发展规模和研究速度。据 Scopus 数据显示,1923 年至 2004 年基因编辑技术的世界年文献发表量均在个位数以下,2005 年 13 篇,2006 年 11 篇。自 2007 年开始,该领域的年文献发表量呈显著上升趋势,因此我们选择 2007 年至 2017 年的年文献发表量作为研究对象。分析 2007—2017 年基因编辑领域每年的发文量可以看出(见图 1),2007 年世界总发文量为 26 篇,2007—2012 年呈线性增长,2012 年文献发表量为 207 篇;2012 年至今该领域发表文献数呈指数型增长,2016 年达 2 153 篇。特别是自 2012 年 M·Jinek 等^[8]在体外率先证实了 Cas9 可以在人工合成的 sgRNA 引导下对靶标 DNA 序列进行特异切割以来,CRISPR 技术以其编辑效率更高、操作更简便、成本更低等相对优势,迅速成为当前基因组编辑技术的主流。该技术已三度入围美国《科学》杂志年度十大突破,更在 2015 年被《科学》评为年度头号突破。与 CRISPR 技术有关的论文数量呈爆发式增长,也是基因编辑技术领域总体论文量自 2012 年起呈指数级增长的关键因素之一。

中国该领域年发文量变化趋势与世界发展水平基本一致,但是中国年发文量快速增加的起始点略晚于世界年发文量。2011 年以前中国基因编辑技术年发文量均低于 10 篇,2012 年为 18 篇,自 2013 年(54 篇)开始形成拐点,随后呈线性增长。2013 年 3 月,云南中科灵长类生物医学重点实验室、中国科学院昆明动物研究所、南京医科大学、南京大学、同济大学等多家单位合作,利用基因组编辑技术 CRISPR/Cas9 和 TALENs 实现猕猴和食蟹猴两个物种的靶向基因修饰,同年 10 月和 11 月分别获得世界首例经过 CRISPR/Cas9 基因靶向修饰及 TALENs 基因靶向修饰的两只小猴。该两项研究成

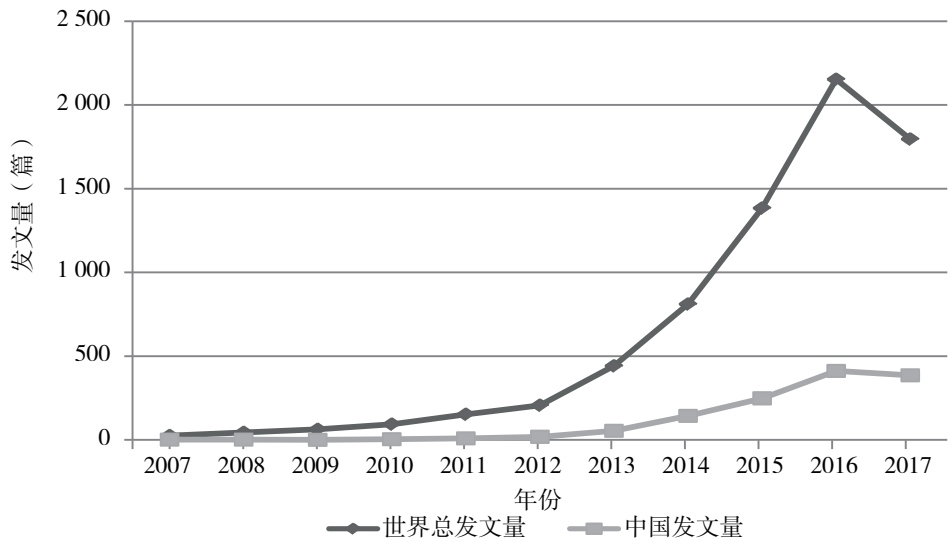


图 1 2007—2017 年年发文量

果分别发表在《细胞》(Cell)杂志和《细胞干细胞》(Cell Stem Cell)杂志,引发全球科学界和媒体的高度关注。

2.2 国家地区及机构分析

对文献发表所属国家地区和机构的分析一定程度上可以展示目前在某一领域不同国家地区和机构的研究水平以及学者对该领域的重视程度。首先对 7 173 篇文献整体发表的国家和地区情况进行分析,结果显示基因编辑技术的研究地区主要集中在美国、欧洲以及东亚地区。进一步整理了文献发表量前 20 名的国家和地区(见表 1),

其中发文量最多的国家是美国,共发表文献 3 360 篇,占总文献发表量的 46.84%,发文数量排名第 2 的是中国,发表文献 1 329 篇,占总文献发表量的 18.53%,发文数排名第 3 的是德国,发表文献 650 篇,占总文献发表量的 9.1%。日本和英国分列第 4 位和第 5 位。

对 7 173 篇文献的发文机构进行统计分析,发表文献数量最多的机构是美国哈佛大学,共发表文献 523 篇,占总文献发表量的 7.3%,排名第 2 的是中国科学院,共发表文献 313 篇,占总文献发表量的 4.4%,排名第 3 的是霍华德·休斯医学研究所,

表 1 发文量排名前 20 的国家 / 地区及发文数量

国家	发表文章数	国家	发表文章数
美国	3 360	丹麦	150
中国	1 329	瑞士	139
德国	650	西班牙	134
日本	552	意大利	126
英国	512	印度	101
法国	352	瑞典	90
加拿大	230	俄罗斯	86
荷兰	227	以色列	73
韩国	223	新加坡	67
澳大利亚	174	巴西	62

共发表文献 267 篇, 占总文献发表量的 3.7%。在世界发文量排名前 20 的机构中, 除了中国科学院发文量位居世界第 2 外, 中国还有中国教育部、复旦大学和北京大学发文量进入排行榜前 20 (见表 2)。

对基因编辑技术排名前 6 位的科研机构发表在世界排名前 10 期刊的文章进行科研质量水平分析, 发现博德研究所 FWCI^① 指数最高 (16.32), 其次是麻省理工学院 (14.76)、哈佛大学 (10.32)、霍华德·休斯医学研究所 (10.19)。中国科学院在科研成果产出量保持高位的同时, 科研质量相对较低, FWCI 指数 3.81; 美国国立卫生研究院不但科研产出相对较低, FWCI 指数也相对较低 (4.94),

与其他几家研究机构相比, 基因编辑技术方面的科研质量相对较低 (见表 3)。

2.3 专利分析

Scopus 的专利数据主要来源于世界知识产权组织 (WIPO)、欧洲专利局 (EPO)、美国专利局 (USPTO)、日本专利局 (JPO) 和英国知识产权局 (IPO.GOV.UK), 共计 2 520 万份专利。根据 Scopus 数据库的专利数据统计分析, 7 173 篇基因编辑技术相关文献共涉及专利 5 400 多份。对 2007—2017 年相关专利统计发现, 2007—2009 年年均专利数不足 100 件, 2010 年增长到 130 件, 2010—2012 年增长缓慢, 自 2013 年开始迅速增长, 从 2013 年的 285 件猛增到 2016

表 2 发文量排名前 20 的机构及发文数量

机构	发表文章数	机构	发表文章数
哈佛大学	523	加州大学旧金山分校	120
中国科学院	313	斯坦福大学	115
霍华德·休斯医学研究所	267	杜克大学	113
麻省理工学院	197	法国国家健康与医学研究中心	102
博德研究所	164	哥本哈根大学	94
美国国立卫生研究院	142	美国劳伦斯伯克力国家实验室	92
伯克利加州大学	139	复旦大学	83
哈佛大学医学院附属麻省总医院	125	北京大学	83
中国教育部	124	京都大学	82
法国国家科学研究院	121	首尔大学	82

表 3 世界排名前 6 位的科研机构在世界排名前 10 期刊的发文量比例及 FWCI 指数

机构名称	文献发表在世界排名前 10 期刊文献数量占总发文量百分比 (%)	FWCI
博德研究所	58.4	16.32
麻省理工学院	58.9	14.76
哈佛大学	44.3	10.32
霍华德·休斯医学研究所	52.2	10.19
美国国立卫生研究院	25.7	4.94
中国科学院	23.1	3.81

① 指域加权引用影响 (Field-Weighted Citation Impact), 通常用以衡量科研质量。一般认为世界平均水平的 FWCI 为 1, 大于 1 说明研究质量高于世界平均水平, 小于 1 则说明低于世界平均水平。

年的 1 690 件,基因编辑技术进入成果转化飞速发展期(见图 2)。

图 3 显示的是世界排名前 6 位的科研机构发表在世界排名前 10 期刊文献的专利引用情况,气泡大小表示文献的专利引用数量(Citing-Patents Count)。由图可见,哈佛大学发表基因编辑技术相关文献数量最多,引用文献的专利数量也最多,共有 339 个专利,位列第一梯队;麻省理工学院和博德研究所在世界排名前 10 期刊发表文献数量占比最高,分别为 58.9% 和 58.4%,且二者在发表文献数量及专利引用数量上都很接近,其发表在世界排名前 10 期刊文献的专利引用数量分别为 264 和 260;霍华德休斯医学研究所专利引用数量为

287,其发表在世界排名前 10 期刊的文献数量占比为 52.2%,与麻省理工学院和博德研究所共同属于第二梯队。中国科学院和美国国立卫生研究院发表在世界排名前 10 期刊的文献量占比和专利引用数量均相对较低,其专利引用数量分别为 57 和 31,位列第三梯队。

2.4 作者分析

统计分析发现,7 173 篇文献中发文数量最多的作者是来自美国北卡罗来纳州立大学生物与营养科学系的 R. Barrangou 教授,参与发表文章 73 篇;其次是美国哈佛与麻省理工学院共建的博德研究所华裔教授 F. Zhang,参与发文量 66 篇;第三名是美国加州大学伯克利分校分子细胞和生物

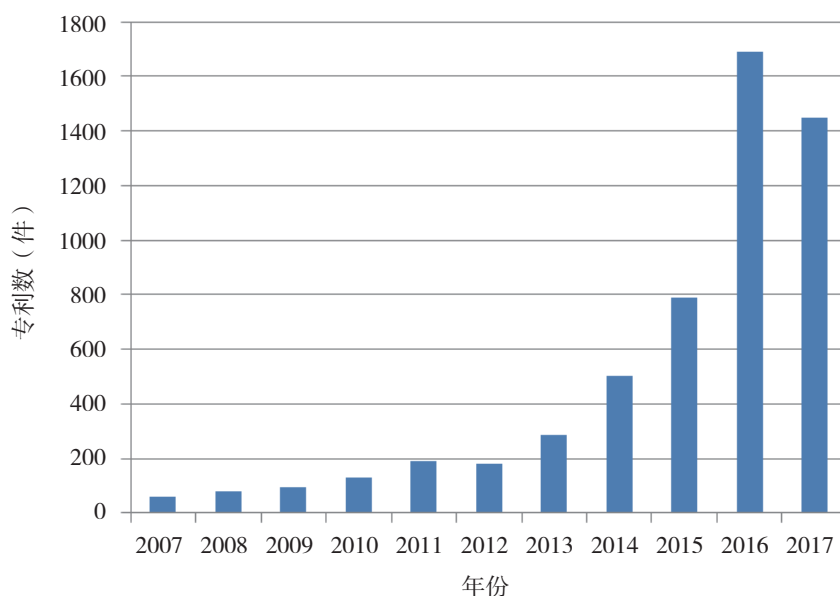


图 2 2007—2017 年历年专利申请量

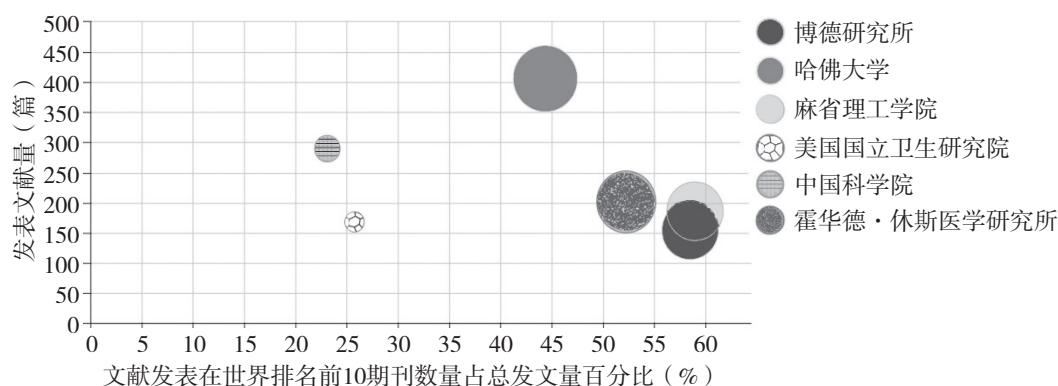


图 3 世界排名前 6 位的科研机构发表在世界排名前 10 期刊文献的专利引用情况

学系女教授 J.A. Doudna, 参与发文量 64 篇。研究发现, 发文量最多的 3 位作者的研究方向均为 CRISPR 基因编辑技术的发明及应用领域。2007 年, R. Barrangou 及其团队首次用实验证明了嗜热链球菌 (*Streptococcus thermophilus*) II 型 CRISPR 系统具有获得性免疫防御功能^[4]; 2012 年 6 月, J.A. Doudna 团队首先在线发表了有关 CRISPR 技术的论文, 标志着 CRISPR 基因编辑技术的诞生^[8]。同年, F. Zhang 团队也发表相关文章, 首次证明 CRISPR 技术能应用于人类细胞的基因组^[9], 并快速获得了 CRISPR 技术的第一个专利。近年来, F. Zhang 团队在提高 CRISPR/Cas9 基因编辑系统的效率方面做了大量研究工作^[10-12]。

为了进一步扩大作者研究范围, 确定核心作者

群, 根据文献计量学的普赖斯定律, 某一领域中核心作者的最低发文数量应满足如下公式:

$$m \cong 0.749 \times \sqrt{n_{\max}}$$

式中, $n_{\max}$ 为最高产的作者的发文数; m 为核心作者的最低发文数。

由此公式可推算出基因编辑技术相关研究的核心作者的最低发文量为:

$$m \cong 0.749 \times \sqrt{73} = 6.399 \approx 6$$

通过公式计算可以发现, 发文在 6 篇及以上的作者为基因编辑技术的核心作者, 共 158 位, 这 158 位作者共发文 2 752 篇, 占总发文数量的 38.37%。由于篇幅原因, 本文整理了发文数量前 20 名的作者及其参与发表的文献数量 (见表 4)。

对发文数量前 10 名作者的历年发表文章

表 4 参与文献数量排名前 20 的作者、发文量及其所属机构

作者姓名	参与文献数	所属研究机构	作者姓名	参与文献数	所属研究机构
R. Barrangou	73	美国北卡罗来纳州立大学	M.C.Holmes	39	美国加州里士满的桑加莫 生物科技公司 (Sangamo Biosciences)
F. Zhang	66	美国博德研究所	T. Cathomen	39	德国柏林夏利特 (Charité) 医学院
J.A. Doudna	64	美国加州大学伯克利分校	C.A. Gersbach	34	美国杜克大学
T. Yamamoto	61	日本广岛大学	L.A. Marraffini	31	美国西北大学
J.S. Kim	58	韩国首尔国立大学	K.S. Makarova	30	美国国立卫生研究院
D.F. Voytas	56	美国明尼苏达大学	S.J.J. Brouns	30	荷兰瓦赫宁根大学
P.D. Gregory	48	美国加州里士满的桑加莫 生物科技公司 (Sangamo Biosciences)	R.A. Garrett	29	丹麦哥本哈根大学
T. Sakuma	47	日本广岛大学	P.C. Fineran	29	新西兰奥塔哥大学
E.V. Koonin	47	美国国立卫生研究院	G. Bao	29	佐治亚理工学院
J.K. Joung	47	美国马萨诸塞州综合医院 (MGH)	B. Wiedenheft	28	美国霍华德·休斯 医学研究所

数量做进一步统计分析发现, 2007 年只有 R. Barrangou、P.D. Gregory 和 D. Philip 在基因编辑技术方向有文章发表, 其他 8 位作者均自 2008 年及其后才开始有相关文章发表。尤其值得注意的是, F. Zhang 教授和 T. Sakuma 教授在基因编辑技术方面的相关文章自 2012 年才开始发表, 随后迅速增

长, 成为该领域的核心人物 (见图 4)。

为了从整体水平把握核心作者的科研能力, 对发文数量前 10 名作者的个人文章总量的历年被引用数量进行了统计分析 (排除所选作者的自引项, 见图 5)。分析可见, E.V. Koonin 研究员的历年文献被引数量最高, 其 h 指数为 152, 也为几位作者

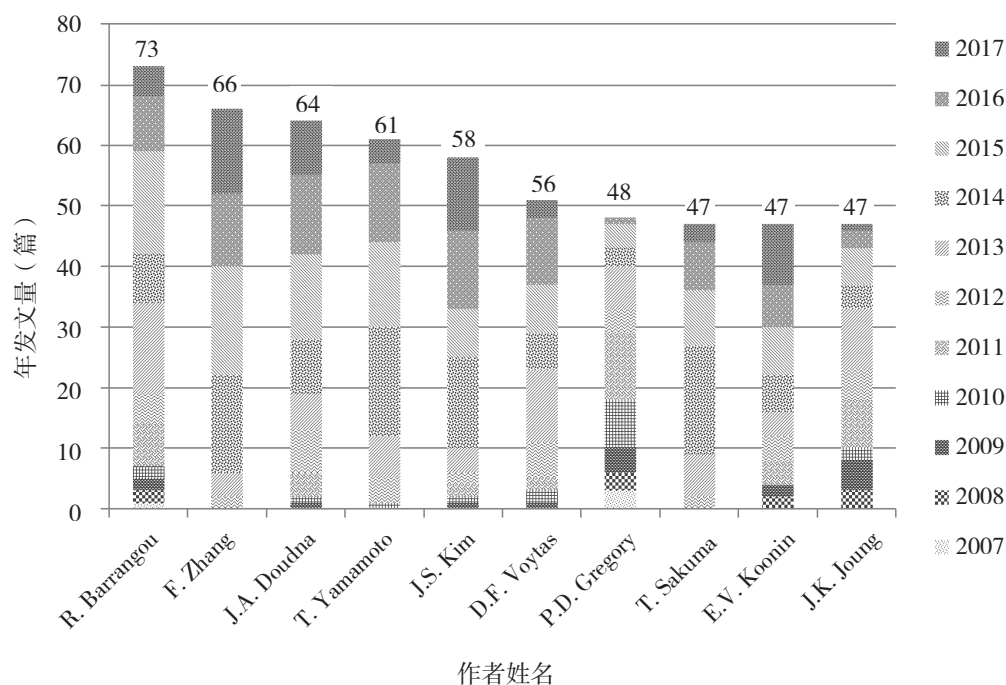


图4 前10名作者年发文量

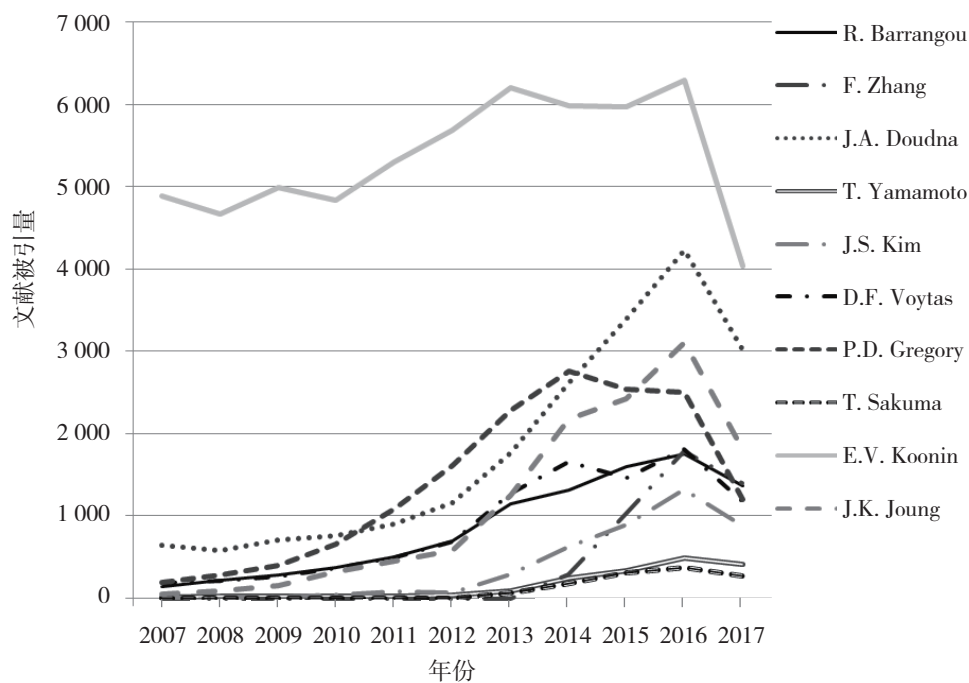


图5 前10名作者文献被引量

中最高。但是其基因编辑技术方向文章数量排名第9,自2008年开始参与发表基因编辑相关文章后,每年发表文章数也相对均衡,由此可见E.V. Koonin科研能力处于国际顶尖水平,但是其研究方向多样,基因编辑技术很可能只是其中一个研

究分枝。J.A. Doudna教授的历年来文献被引量也非常显著,其 h 指数为81,仅低于E.V. Koonin。2009年之前J.A. Doudna的文献被引数量无显著变化,自2009年开始发表基因编辑技术相关文章后,其文献被引量变化趋势与基因编辑技术文献年发

表数量趋势一致,呈指数性增加,表明自2009年以后 J.A. Doudna 的主要研究方向为基因编辑技术并在该领域成为顶级核心作者。F. Zhang 教授2012年之前的文献被引量均为0,自2013年开始呈爆发式增长,结合其文章发表情况也是从2012年开始,说明 F. Zhang 的研究领域集中在基因编辑技术,并且在短时间内成为该领域的核心人物,发展潜力巨大。

为进一步分析中国在基因编辑技术领域的研究现状,统计出158位核心作者中的中国学者及其研究机构(见表5),其中发文量最多的分别是上海科技大学黄行许教授、吉林大学赖良学教授以及中国科学院遗传与发育生物学研究所高彩霞研究员。先后在南京大学模式动物研究所、上海科技大学任职的黄行许,是近年来基因组编辑领域

的风云人物之一。近期,黄行许教授课题组接连在CRISPR/Cas9研究中取得突破性成果,相关论文发表在Nature Methods、Cell Research、Scientific Reports 和 Human Molecular Genetics 等权威期刊。在农作物育种方面,高彩霞研究组在小麦中率先建立了基于CRISPR/Cas9瞬时表达的基因组编辑体系并致力于对该体系进行不断改良;朱健康院士团队成功利用改良的CRISPR系统精确编辑水稻基因组。

2.5 期刊分析

通过对数据进行整理发现,检索的7173篇文献共发表在1287种不同的期刊当中。文献计量学中的布拉德福定律指出,如果将科技期刊按其刊载某学科专业论文的数量多少,以递减顺序排列,那么可以把期刊分为专门面对这个学科的核心区、相

表5 中国核心作者发文量及其发文机构

作者姓名	参与文献数(篇)	研究机构
黄行许	22	上海科技大学、南京大学
赖良学	20	中国科学院广州生物医药与健康研究院、吉林大学
高彩霞	15	中国科学院遗传与发育生物学研究所
周建奎	13	上海科技大学
张智英	12	西北农林科技大学
李占军	10	吉林大学
朱健康	10	中国科学院上海生命科学研究院
李大力	9	华东师范大学

关区、非相关区。这3个区中的期刊刊载文章数量相等,此时核心区、相关区、非相关区期刊的数量成 $1:n:n^2$ 的关系。

统计发现,发表文献数量在46篇及以上的期刊有24种,共发表文献2439篇,约占总文献数量的32.74%;发表文献数量在8~43篇的期刊总共有139种,共发表文献2453篇,约占总文献量的34.20%;发表文献数量在1~7篇的期刊1124种,共发表文献2280篇,约占总文献数量的31.79%。核心区期刊、相关区期刊、非相关区期刊的数量大约成 $1:6:36$ 的关系(见表6)。

对核心期刊及其发文数的统计发现,发文数最多的期刊前3名为Scientific Reports、PLoS ONE 和

Nucleic Acids Research, 对应发文数分别为377、314和211篇。核心期刊中被引次数(Cite Score)为10以上的期刊有7种,总发文量579篇,占核心期刊总发文数的23.74%;其中CNS(Cell, Nature, Science)的总发文量为248,占核心期刊总发文数的10.17%。顶级核心期刊数量及发文量表明该领域的科研质量,凸显了研究的创新性及突破性意义(见表7)。

2.6 学科分布分析

目前基因编辑技术涉及的研究覆盖近30个学科领域,因篇幅所限,主要节选了发表文章数量最多的前20个学科领域(见图6),由于学科间有交叉,因此总百分数大于1。目前使用基因编辑进行研究的主要方向有生物化学、遗传和分子

表 6 核心区期刊、相关区期刊、非相关区期刊数及发表文献总数

期刊分类	期刊数(种)	发表文献总数(篇)
核心区期刊	24	2 439
相关区期刊	139	2 453
非相关区期刊	1 124	2 280

表 7 核心期刊发文数

刊名	文献数	被引次数 (2016)	刊名	文献数	被引次数 (2016)
Scientific Reports	377	4.63	Journal of Biological Chemistry	69	4.17
PLoS ONE	314	3.11	Molecular Therapy	65	5.60
Nucleic Acids Research	211	9.28	ACS Synthetic Biology	57	4.70
PNAS	152	8.56	BMC Genomics	57	4.05
Methods in Molecular Biology	120	—	Science	56	14.39
Nature	111	13.33	Nature Methods	54	11.58
Nature Communications	110	11.80	Biochemical and Biophysical Research Communications	53	2.51
Nature Biotechnology	93	13.16	eLife	52	5.41
Cell	81	22.79	Frontiers in Microbiology	47	4.16
Oncotarget	75	4.73	Genetics	47	4.70
Molecular Cell	74	12.18	PLoS Genetics	46	5.93
Cell Reports	72	8.40	RNA Biology	46	3.37

生物学,发文数量 4 900 篇,占总文献发表量的 68.31%;医学,发文数量 2 550 篇,占总文献发表量的 35.55%;免疫和微生物学,发文数量 1 155 篇,占总文献发表量的 16.10%;农业与生物科学,发文数量 1 151 篇,占总文献发表量的 16.04%。科研质量方面,FWCI 指数排名最高的领域是分别是工程学、化学工程和多学科交叉科学;而发文量最多的四大学科(生物化学、遗传和分子生物学,医学,免疫和微生物学,农业与生物科学)的 FWCI 指数相对均衡,平均为 3 左右。

2.7 关键词共现分析

利用 VOSviewer 软件对 7 173 篇文献题目和摘要基于同现词网络构建关键词术语地图,发现基因编辑技术涉及的研究被聚类成典型的几个方向(见图 7),其中左下方聚类包含 401 个术语,出现的

高频关键词有锌指核酸酶(zinc finger nuclease)、转录激活因子样效应因子核酸酶(transcription activator/talen)、动物模型(animal model)、植物(plant)、多功能干细胞(pluripotent stem cell)等;左上方聚类包含 287 个术语,出现的高频关键词有治疗(treatment/therapy)、响应(response)、癌症(cancer)、抑制(inhibition)、受体(receptor)、互作(interaction)等;右侧聚类包含 267 个术语,出现的高频关键词有重复(repeat)、CRISPR/Cas 系统(crisprcasystem/cas)病毒(virus)、细菌(bacterium)、菌株(strain)、间区序列(spacer)、多样化(diversity)等。图 8 所示的研究热点主题密度图中颜色越深表明词频出现概率越高,越趋向于研究热点,对深色区域的关键词进行综合分析,得出的主要研究领域有:(1)ZFN 和 TALEN 两种

基因编辑技术在植物育种和动物模型方面的应用；
 (2) CRISPRs-Cas9 在癌症靶向治疗方面的应用；
 (3) 利用 CRISPR 基因编辑技术控制蛋白表达以治疗与蛋白相关的疾病；(4) 对 CRISPRs-Cas 细菌免疫系统自身的不断完善研究，如降低脱靶率、提高 CRISPR 适应的重复精度等。

图 9 的时间热度地图展现了基因编辑技术在不同主题中的演变情况。由图可见，文章类型从 2013 年单纯的技术性文章，逐渐向医学、动植物育种等应用性文章转变，这也符合新兴技术从基础理论研究到应用转化研究的演变趋势。基因编辑技术未来的应用方向主要集中在癌症 / 肿瘤的靶向治疗、药物研发、蛋白调控、植物育种等方向。

3 讨论

3.1 基因编辑技术整体发展态势迅猛，中国与世界顶尖水平仍有差距

对 2007—2017 年 9 月的文献进行分析可知，基因编辑技术自 2012 年 CRISPR 技术正式诞生以来发展迅速。从文献量上看，近 5 年来基因编辑技术相关文献发表数量不断提高，预计未来几年还将有大幅提升。专利引用情况也是自 2013 年起有了显著增加，2016 年引用文献专利数已达 1 690 篇，截至 2017 年 9 月份已达 1 400 余篇。特别是博德研究所 F. Zhang 团队与加州大学伯克利分校 J.A. Doudna 团队关于 CRISPR/Cas9 技术之间的专利之争，更加凸显了在 CRISPR/Cas9 技术专利申请方面

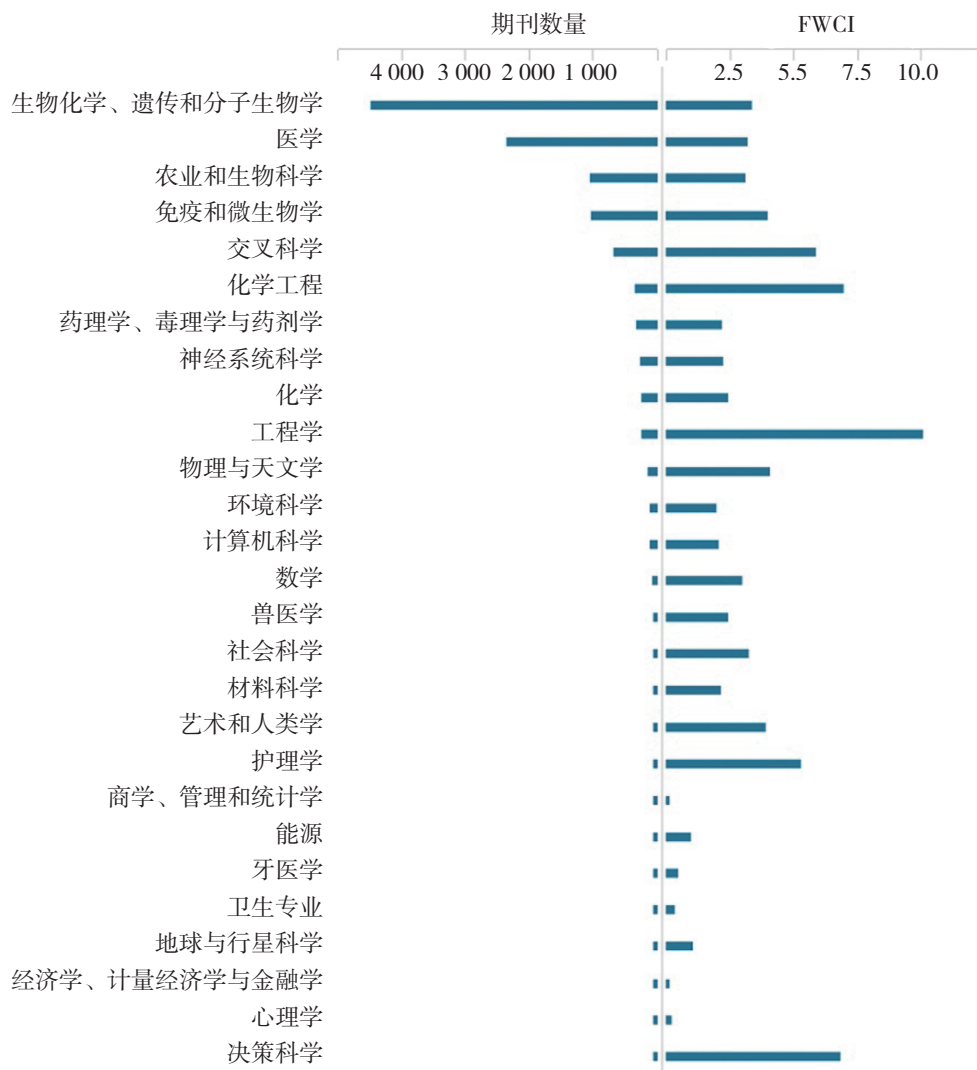


图 6 学科分布及科研质量分析

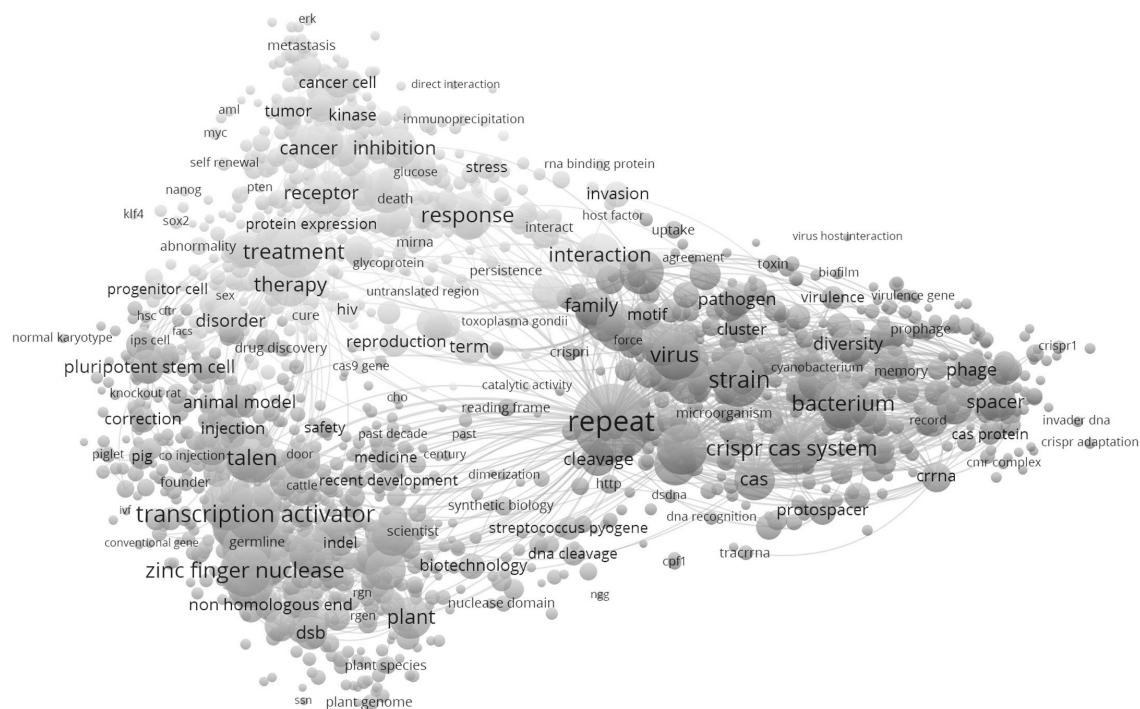


图 7 VOSviewer 构建基因编辑技术关键词术语地图

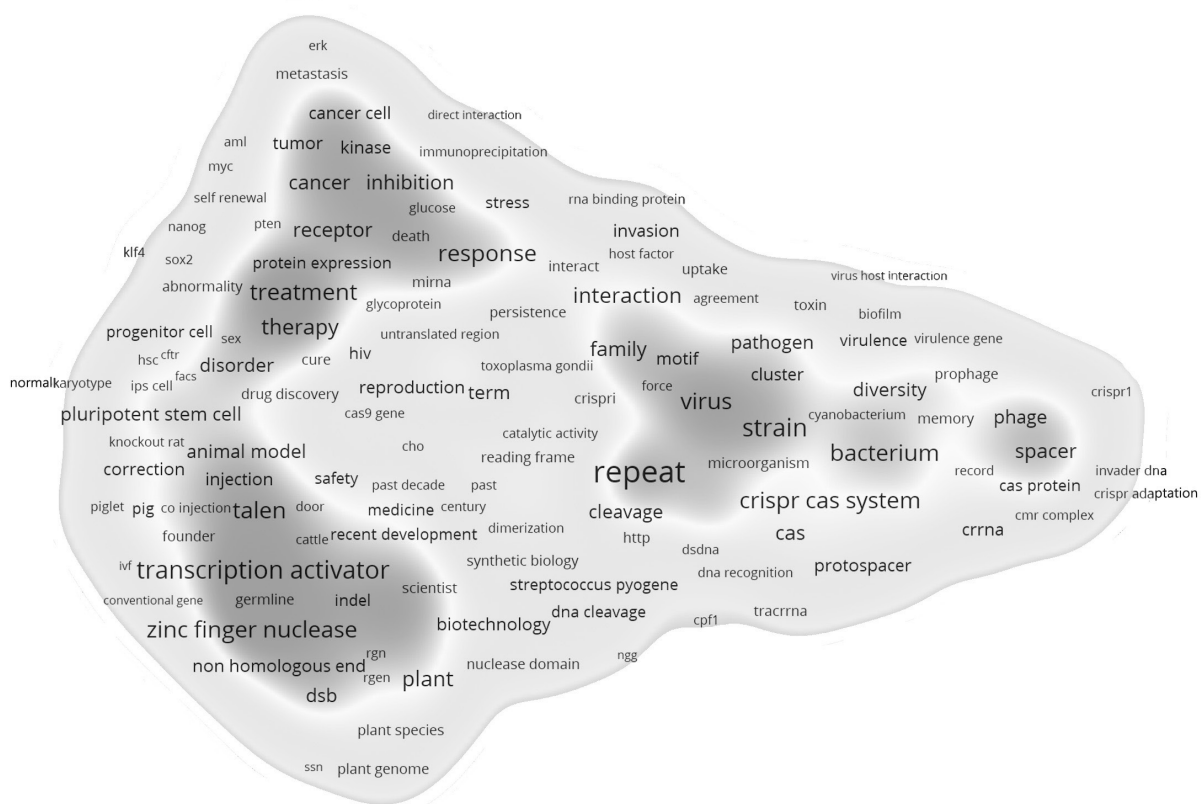


图 8 VOSviewer 构建基因编辑技术研究热点主题密度图

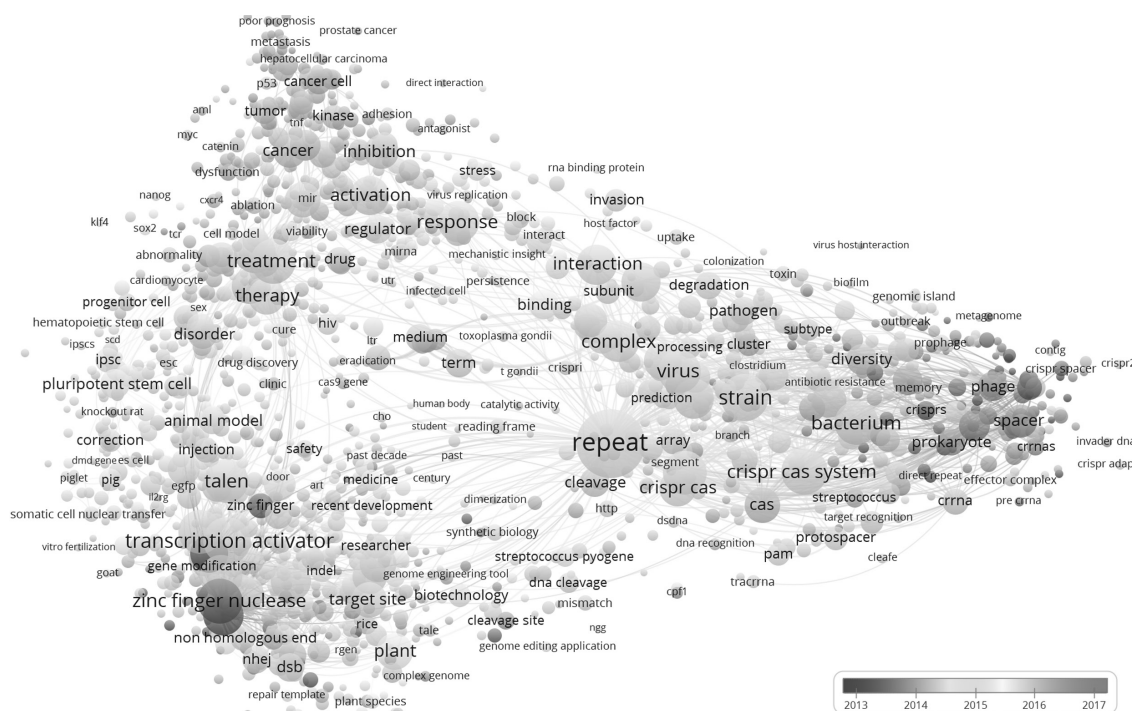


图 9 VOSviewer 构建基因编辑技术研究趋势变化图

的白热化竞争。从国家 / 地区、机构的发文量排名可见, 目前基因编辑技术的研究与应用主要集中在美国、中国、欧洲以及日本和韩国。中国的发文量仅次于美国, 已在该领域排名世界第二。但是中国论文影响力和专利价值方面和国际顶尖机构相比仍有较大差距, 应继续加强布局和研究。经济影响方面, 美国哈佛大学、麻省理工学院、博德研究所等机构专利引用数量位居世界前列。

3.2 基因编辑技术已初步形成作者集群，中国研究人员的研究方向分布广泛

从研究的作者群体上看,目前基因编辑技术领域已初步形成作者集群,主要集中在一些世界顶级的研究机构和大学。对 158 位核心作者的隶属机构和发文情况的分析表明,美国学者在发文数量和质量上均有显著优势。特别是以 CRISPR/Cas9 技术出现为代表的一批研究人员,如 F. Zhang、T. Yamamoto、T. Sakuma 等,在短短几年时间内迅速成为该领域的核心人物。中国学者在该领域取得了一系列重大突破,对核心作者中的 8 位中国学者的分析表明,其研究方向集中在基因编辑技术在人类医学、动物和植物各个方面的具体应用,为该技术的发展完善及应用做出了重大贡献。

3.3 基因编辑技术研究方向广泛，逐渐从纯技术研究转向各个领域的实际应用

目前基因编辑技术方面的核心区期刊 24 本, 共发表文章 2 439 篇, 其中 CNS 发文量占比 10.17%, 其所在的期刊和论文影响力水平已经达到顶级。研究方向方面, 目前的文献研究覆盖了近 30 个学科, 文献量分布最多的几个学科领域是生物化学、遗传和分子生物学, 医学, 免疫和微生物学, 农业与生物科学, 主要围绕生命科学领域的方方面面。研究方向已经从纯技术研究向各个领域的应用转化, 将创造巨大的经济价值和社会价值。最新的前沿热点研究主题有癌症 / 肿瘤的靶向治疗、药物研发、蛋白调控、植物育种等。例如, 基因组编辑技术在农作物育种方面已经显示出巨大的潜力和优势, 主要体现在通过该技术能够创制出数量性状的优异等位变异, 这是以往技术难以实现的。目前国内多个研究团队在不同作物上都建立了基因编辑技术, 为广泛利用这类技术对作物品种进行定向改良奠定了基础。

4 结论

生命科学关系人类的生存、健康和可持续发

展,是21世纪最重要的学科领域之一。基因编辑技术的快速发展,使对基因的定向修饰更为简便有效,对靶向治疗、动植物遗传育种、新蛋白制造和药物研发等领域造成了颠覆性创新,大大扩展了从基础生物学到生物技术和医学的应用范围。加大对该技术领域的投入研究,对提高我国人民健康和生活水平、改善环境质量、确保粮食生产和安全,将发挥重要作用。■

参考文献:

- [1] Patrick D Hsu, Eric S Lander, Feng Zhang. Development and Applications of CRISPR-Cas9 for Genome Engineering[J]. Cell, 2014 (157): 1 262-1 278.
- [2] Miller J C, Holmes M C, Wang J, et al. An improved zinc-finger nuclease architecture for highly specific genome editing[J]. Nature. Biotechnology, 2007 (25): 778-785.
- [3] Christian M, Cermak T, Doyle E L, et al. Targeting DNA double-strand breaks with TAL effector nucleases[J]. Genetics, 2010 (186): 757-761.
- [4] Barrangou R, Fremaux C, Deveau H, et al. CRISPR provides acquire d resistance against viruses in prokaryotes [J]. Science, 2007 (315): 1 709-1 712.
- [5] 邱均平. 文献计量学[M]. 北京: 科学技术文献出版社, 1988: 15-24.
- [6] 刘彬, 邓秀新. 基于文献计量的园艺学基础研究发展状况分析[J]. 中国农业科学, 2015 (48): 3 504-3 514.
- [7] 高凯. 文献计量分析软件 VOSviewer 的应用研究[J]. 科技情报开发与经济 [J]. 2015 (25): 95-98.
- [8] 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.
- [9] Cong L, Ran, F A, Cox D, et al. Multiplex genome engineering using CRISPR/Cas systems[J]. Science 2013 (339): 819-823.
- [10] Yamano T, Nishimasu H, Zetsche B, et al. Crystal structure of Cpf1 in complex with guide RNA and target DNA[J]. Cell, 2016 (165): 949-962.
- [11] Cox D B T, Gootenberg J S, Abudayyeh O O, et al. RNA editing with CRISPR-Cas13[J]. Science, 2017 (358): 1 019-1 027.
- [12] Gootenberg J S, Abudayyeh O O, Kellner M J, et al. Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a and Csm6[J]. Science, 2018 (360): 439-444.

Analysis of International Development Trend of Gene Editing Technology Based on Bibliometrics

CAO Xue-wei, GAO Xiao-song, CHEN Rui

(National Academy of Innovation Strategy, China Association for science and Technology Beijing 100012)

Abstract: In recent years, a series of programmable nuclease-based genome editing technologies have been developed, enabling targeted and efficient modification of a variety of eukaryotic and particularly mammalian species. In order to know the development status quo of this field in China, this paper analyzes the international research progress of gene editing technology from 2007 to 2017 with the method of bibliometrics. Taking Scopus as data source, this paper analyzes the total volume of papers during this period, core authors, core journals, research countries and institutes, major subject distribution, economic impact (patent analysis) and research frontier and hotspot. The results of this analysis can help us know the trend of gene editor technology better, and show the research level of China in world. It provides a global perspective to grasp the opportunities and challenges of emerging industry innovation.

Key words: gene editor; bibliometrics; core authors; research institutes; patent analysis; research frontier and hotspot