

# 基于知识图谱的国内高血压研究现状与趋势<sup>\*</sup>

王 洁 乔海波 彭 岩 薛锦亿

(首都师范大学管理学院 北京 100089)

〔摘要〕 以中国知网中高血压主题文献为研究对象,采用知识图谱分析国内高血压领域中研究机构分布、作者之间形成的团队模式及合作程度以及目前研究热点、历年研究内容变化和未来研究重点,全面展示国内高血压研究现状与趋势。

〔关键词〕 高血压; CiteSpace; 知识图谱; 可视化

〔中图分类号〕 R-056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2017.12.012

**Domestic Hypertension Study Status and Trend Based on Knowledge Map** WANG Jie, QIAO Hai-bo, PENG Yan, XUE Jin-yi, School of Management, Capital Normal University, Beijing 100089, China

〔Abstract〕 The paper takes hypertension literature on CNKI as its study subject and adopts knowledge map to analyze research institution distribution, team modes and cooperation degrees formed among authors, transformation of study contents over the years, current study hotspots and future emphases for study in domestic hypertension field, giving a full picture of domestic status and trend of hypertension.

〔Keywords〕 Hypertension; CiteSpace; Knowledge map; Visualization

## 1 引言

随着城镇化、人口老龄化和疾病谱的改变,以及社会经济发展模式的转型,高血压等慢性非传染性疾病已成为我国重要的公共卫生问题<sup>[1]</sup>。高血压病是以血压升高为主要特征的疾病,它不仅表现为血流动力学的异常改变,同时伴有糖、脂质代谢异

常,以及心、脑、肾等靶器官的损坏<sup>[2]</sup>。2016 年 12 月孙宁玲教授“中国高血压进展与管理模式探索”的报告<sup>[3]</sup>显示,虽然目前我国居民高血压知晓率、服药率与控制率有显著的进步,但高血压患病率仍逐年上升。2016 年发表在《美国医学会杂志》上的关于我国高血压队列研究<sup>[4]</sup>的调查显示,我国高血压患病率为 32.5%,且高血压患病率随年龄逐步增高。

由于高血压疾病的广受关注,我国学者从不同角度对其开展研究,取得一定的科研成果。以中国知网数据库(CNKI)为数据来源统计显示,在 1992-2016 年期间,我国学者发表高血压相关科研文献高达 40 000 余篇,其中在 1992-2005 年期间,每年的发文量均在 1 000 篇以上(除 1997 年),

〔修回日期〕 2017-09-05

〔作者简介〕 王洁,博士,副教授,硕士生导师,发表论文 50 余篇。

〔基金项目〕 国家科技基础性工作专项“中国主要少数民族体质人类学系统表型特征调查”(项目编号:2015FY111700-6)。

2006 年至今每年的发文量更是翻一番。这些科研文献虽然主要集中在医疗卫生科技领域,但也涉及多个不同的学科;研究内容错综复杂,涉及高血压防治、患病因素、心血管病、临床药物研究等众多方面内容。面对数量庞大,研究内容众多且相互交叉的科研文献,研究者们期望能对现有文献进行梳理,以期后续研究提供重要参考。本文采用文献计量和可视化的研究方法,运用 CiteSpace III 分析工具,以 CNKI 中检索得到与高血压研究相关的文献为数据源,梳理我国高血压领域的研究成果,以知识图谱的形式展现高血压的研究现状、研究热点和发展趋势,以期对我国高血压领域的进一步研究提供有益参考。

## 2 资料与方法

科学知识图谱不仅能揭示知识来源及其发展规律,并且以图形表达相关领域知识结构关系和演进规律。本研究所使用的工具为 CiteSpace III 可视化软件,是由美国 Drexel 大学陈超美教授开发的基于 JAVA 平台的一款运用十分广泛,多元、分时、动态的可视化软件<sup>[5]</sup>。其主要功能包括关键词共现分析、机构合作分析、文献被引分析,通过聚类视图和时区视图可以快捷清晰地绘制出研究热点、动向、前沿与发展趋势。本文研究数据收集于 CNKI 数据库,发文年限设置为 2006 - 2016 年,检索主题设置为“高血压”,学科领域设置为“心血管系统疾病”、“临床医学”、“外科学”等医疗卫生科技研究领域,来源期刊限定为 SCI 源期刊、EI 源期刊、核心期刊、CSSCI,共获得数据 24 719 条。文献的被引次数能够很好地说明文献在相应领域的重要性,因此本文对以上检索数据进行被引排序,取其排名靠前的文献,得到有效数据 2 000 条,数据截止日期为 2016 年 12 月 31 日。将上述文献数据导出,导出的数据包括标题、作者、机构、关键词、发文时间等属性,命名为 CiteSpace 可识别的文件名称,经 CiteSpace 自带数据转换器处理,供可视化分析。

## 3 结果与分析

### 3.1 时间分布 (图 1)

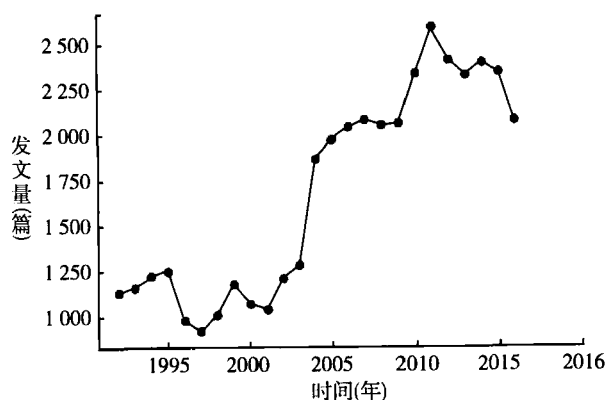


图 1 高血压文献研究年度发文数量变化趋势

从图 1 可看出,1992 - 2016 年期间,除 1997 年外我国高血压研究每年的发文量均在 1 000 篇以上,表明该领域的研究处于上升时期。在数量上整体呈增长趋势,分别在 2005 年和 2011 - 2012 年出现两次快速增长点,说明在这两段时间内出现了新的研究热点或突破口。在 2012 年之后,我国高血压研究发文量开始有所减缓,但仍处于一个较高的水平,这说明我国近 5 年内对高血压研究领域的关注已经相对稳定,此时需要对以往的研究热点进行总结,发掘新的突破口,为进一步深入研究做准备。另外,由图 1 可见,我国高血压研究年度发文量的变化趋势可大致分为两个时期,分别是 1992 - 2005 年和 2006 - 2016 年。前一时期的年均发文量在 1 000 ~ 2 000 篇以内,后一时期年均发文量在 2 000 篇以上,数量上翻了一番,因此后一时段的文献是本文知识图谱分析的主要数据来源。

### 3.2 空间分布

3.2.1 机构分布 设置文献学科为医疗卫生科技学科领域,考察我国高血压研究的核心机构或团体。中国知网中对所有高血压主题下科学文献的机构统计数据显示,在 CNKI 中发表过高血压相关论文的研究机构众多,主要是医院团体,分别为解放军总医院 (2 550 篇)、中国医学科学院阜外心血管

病医院（567 篇）、北京大学人民医院（403 篇）、南京医科大学第一附属医院（292 篇）、首都医科大学附属北京安贞医院（288 篇）和北京大学第一医院（278 篇）等。其中解放军总医院更是以 2 550 篇的发文量位居榜首，说明该研究单位在高血压研究领域具有极强的科研实力和影响力，成为国内高血压研究领域的核心研究机构，此外中国医学科学院阜外心血管病医院和北京大学人民医院等机构也拥有相当强的实力。设置文献数据年限为 2006 - 2016 年，考察不同机构之间的合作情况。在 CiteSpace 中设置 Article Labeling Threshold = 4，生成的高血压研究机构合作图谱，见图 2。每个节点（圆点）代表一个机构，节点的大小代表机构的发文量，标签字号大小代表中心性，节点之间的连线表示合作关系，节点的颜色表示节点初现的年份，由冷色逐渐变暖色，蓝色代表较早的年份，红色代表最近的年份<sup>[6]</sup>。结果显示， $N = 48$ ， $E = 20$ ，Density = 0.0177，图谱密度偏低，说明高血压国内该研究领域的团体比较分散，零散的机构偏多，相互之间的合作还有待加强。从中心性角度观察，几乎所有节点的中心性都为 0，说明没有对其他节点产生影响。而中心性大于 0 的机构只有山东中医药大学附属医院（0.01），说明该机构的研究与其他的机构产生影响，绝大多数的机构影响力仍处在较低的水平，表明机构之间的合作程度还远远不够。考察不同机构在不同时间窗中的研究情况，设置时间片为 1 年，文献数据年限为 2006 - 2016 年，得到高血压机构研究时序图谱（Threshold = 5），见图 3。上方年代颜色对应节点年代的颜色，下方为年代时序<sup>[7]</sup>。结果显示有关高血压的研究 2006 - 2016 年从未间断，每年都有不同的研究机构进入，但高峰期在 2015 年以前，高血压研究的机构已形成规模。其中北京大学人民医院心脏中心和中国医学科学院早在 2007 年就在相关期刊上发表了研究成果，文献量相对较多且数量相当，但前者在图中呈现的整体影响力并不大，二者同山东中医药大学、国家心血管病中心等机构的发文量都不相上下。另外，CiteSpace 中的研究机构突发性监测结果显示，在 2014 年河北联合大学附属开滦总医院心内科、西安

交通大学医学院第一附属医院心内科和国家心血管病中心加入到高血压研究领域，介入该领域研究的迹象明显，且一直参与高血压的研究，可能与 2014 年出现新的研究热点有关。



图 2 高血压研究机构合作图谱



图 3 高血压研究机构时序图谱

3.2.2 作者分布 考察我国高血压研究作者的分布情况。从检索到的全部文献中，分析得到发文量排名前 10 的高产作者分别是：孙宁玲、胡大一、黄如训、李南方、吴寿岭、廖玉华、谢良地、曾春雨、苏海和刘国仗，他们的发文量均在 45 篇以上，数量极多。其中，北京大学人民医院孙宁玲教授更是以 169 篇的发文量位居榜首，表明该作者在高血压研究方面很有建树，具有较高的权威性，是这一领域的专家。考察作者的合作情况，文献数据年限为 2006 - 2016 年。设置 Article Labeling Threshold = 4，生成高血压研究作者的合作图谱。标签字号大小代表中心性，边描述合作情况。结果出现作者 146 名，作者之间的合作有 245 个，说明作者间存

在着较好地合作研究<sup>[8]</sup>；另一方面，作者合作图谱中存在许多子网，说明存在多个研究的小团体，团体内部的合作较多，但团体间的联系与合作较少。进一步调视图谱，得到高血压研究作者合作图谱最大子网络，见图 4。较深较粗的连线颜色表明作者间相互合作时间较早，次数较多。图谱显示，目前已经形成以孙宁玲、吴寿岭、王文等为关键节点的一个核心研究网络，规模最大，作者团体之间的合作呈现出一核心、多组团的研究格局。

## 4 研究内容

### 4.1 关键词共现网络图谱<sup>[9]</sup>

4.1.1 关键词共现 考察高血压研究主题情况，统计关键词出现的频次与高中心词。设置 Threshold = 39，其他属性同上，得到关键词共现图谱，见图 5。关键词共现图谱中包含 162 个节点，436 条连线，说明关键词之间的联系较强。图中的节点代表关键词，节点越大，代表关键词词频越大。两个节点之间的连线表示两个关键词共现，线的粗细代表共现程度<sup>[10]</sup>。结果显示关键词共现网络结构比较紧密，结合图 6 中的关键词突发性监测结果可知，在 2010 - 2011 年中，先后有社区卫生服务、治疗率、代谢综合症、肥胖、依从性、健康教育等关键词突现，成为这一阶段的研究热点。合并“高血压”和“高血压病”等意义相同的关键词后，可以看出“危险因素”、“原发性高血压”、“患病率”、“糖尿病”、“影响因素”等成为高血压研究领域的高频热点词汇。而且这些关键词都在图中表现出较高的中心性，由此可初步判断“危险因素”、“患病率”、“糖尿病”、“影响因素”是研究热点。

4.1.2 各年度关键词共现 由于关键词数目较多，在关键词的共现图谱上并不能很好地观察出每年的研究热点。为得到这些信息，在可视化中选择“显示时间切片网络 (Show Networks by Time Slices)”，其他参数不变，得到了 2006 - 2016 年各年的关键词知识图谱。2006 年的研究热点主要是肥胖症、社区护理、健康教育、农村、老年高血压等关键词；2007 年的研究热点基本上和上一年一样，

但侧重于生活质量、老年高血压和原发性高血压 3 方面的研究；2008 年出现研究热点爆发的现象，主要包括基因多态性、高血压前期、治疗率、体征指数、服药依从性、影响因素等关键词，可见从 2008 年开始对高血压进行基因和药物控制的研究；2009 年的研究热点主要是与高血压有关的一系列病症研究，有血脂异常、动脉粥样硬化、高尿酸血症、心力衰竭等病症。此外 2009 年对高血压的研究延续上一年的研究，在纵向研究领域上更进一步。2010 年对高血压的研究有所减缓，大多是对药物的研究，但开始出现对证候的研究；2011 年的研究则是在前几年的研究热点上进行更深层次的研究，内容多样，但仍然集中在对脑出血、冠心病、糖尿病等病理和药物预防、社区护理两方面；2012 年的研究主要在上一年基础上新增流行病学研究和妊娠期高血压疾病两方面的研究，推测这一年高血压的患病人群急剧增加，开始成为常见病的一种；2013 年基本延续上一年的研究热点，但也有学者研究中医疗法；2014 年的突现的研究热点是高血压的危险因素和影响因素两大方面，而且还涉及了数据挖掘和社区的诊疗、预防两个新的领域；2015 年和 2016 年的研究重点基本上是前几年已经出现的研究内容，图 7 展现 2016 年的关键词共现情况。图中发现出现一些不同的内容，如 2015 年关键词有证候、证素和相关性等研究，2016 年出现了中医症型、性别、应激性、队列研究等关键词出现。可见目前在这几方面的研究还处在起步阶段，预计将是未来该领域研究的重点内容。

### 4.2 研究前沿关键词

4.2.1 关键词聚类 研究前沿是科学研究中最先进、最有发展潜力的研究主题或研究趋势，对把握某一研究领域最新研究趋势起到关键作用。借助 CiteSpace 对关键词聚类后，按照 LLR 命名规则以 Keyword 命名<sup>[7]</sup>，得到关键词聚类图谱，见图 8。该图由 284 个节点和 1 018 条连线组成，模板值 (Modularity Q, Q 值) 为 0.513，而模块值在 0.4 ~ 0.8 之间的图谱通常会形成少量自然的网络聚类，符合分析要求。平均轮廓值 (Mean Silhouette, S 值) 为 0.4892，说明图谱聚类结果比较合理<sup>[11]</sup>。结果显示图谱中得到 15 个特征聚类，每个聚类中

包括规模、轮廓值、平均发表年限和类内主要中心词等主要特征，而且每个聚类代表着高血压研究领域的一个研究方向。有关高血压研究关键词共有 15 个聚类，而其中研究已初具规模的只有 7 个，聚类规模大小已达到 20 左右，结合每个聚类的平均发表年限可知，这些领域的研究成果显著，早已在 2010 年左右就达到一定规模，估计领域内的相关问题已经形成共识。

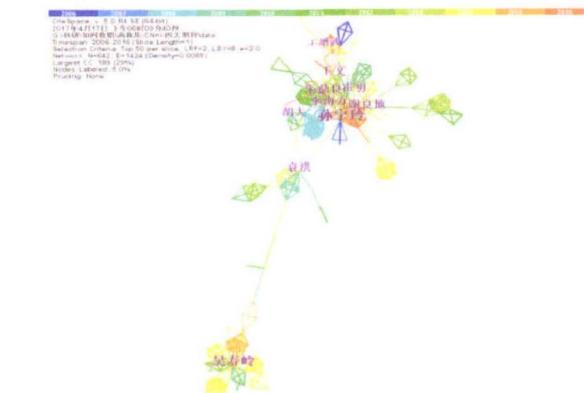


图 4 作者合作图谱最大子网络

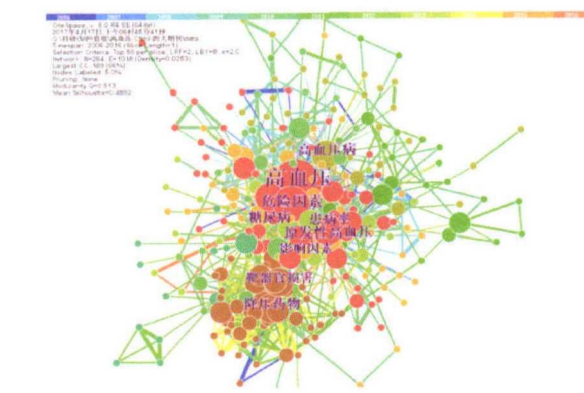


图 5 关键词共现图谱

| 关键词      | 年代   | 力度     | 开始   | 结束   | 2006 - 2016 |
|----------|------|--------|------|------|-------------|
| 一氧化氮     | 2006 | 2.5851 | 2006 | 2007 |             |
| 原发性高血压   | 2006 | 3.0723 | 2007 | 2009 |             |
| 社区卫生服务   | 2006 | 2.9468 | 2008 | 2010 |             |
| 治疗率      | 2006 | 2.5876 | 2008 | 2010 |             |
| 代谢综合征    | 2006 | 2.9788 | 2008 | 2011 |             |
| 肥胖       | 2006 | 5.203  | 2009 | 2012 |             |
| 动态动脉硬化指数 | 2006 | 2.7775 | 2010 | 2011 |             |
| 依从性      | 2006 | 3.7065 | 2010 | 2011 |             |
| 基因多态性    | 2006 | 3.0644 | 2010 | 2012 |             |
| 动态血压     | 2006 | 2.9184 | 2010 | 2012 |             |
| 健康教育     | 2006 | 3.3011 | 2010 | 2012 |             |
| 左心室肥厚    | 2006 | 2.9552 | 2011 | 2012 |             |
| 噻嗪类利尿剂   | 2006 | 3.0467 | 2012 | 2014 |             |
| 难治性高血压   | 2006 | 5.6855 | 2012 | 2014 |             |
| 高血压指南    | 2006 | 2.8502 | 2013 | 2014 |             |
| 诊室血压     | 2006 | 4.225  | 2013 | 2016 |             |
| 血压监测     | 2006 | 3.5658 | 2013 | 2014 |             |
| 动态血压监测   | 2006 | 4.3442 | 2013 | 2014 |             |

图 6 关键词突发性监测



图 7 2016 年关键词知识图谱

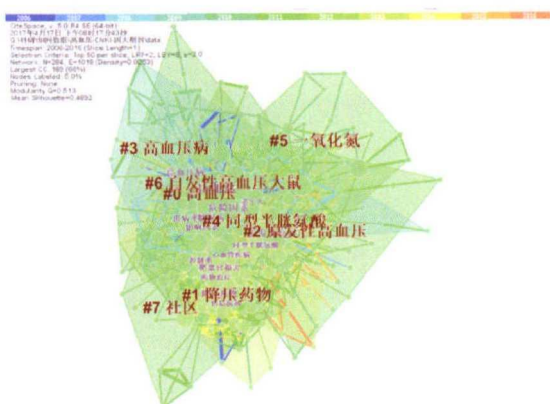


图 8 高血压研究关键词聚类图谱

4.2.2 聚类中心词分布 考察每个聚类的中心关键词时，按规模大小将其分为两类：大规模聚类结果和小规模聚类结果。将第一类聚类的中心词按照规模大小和相应的词频整理后，见图 9。

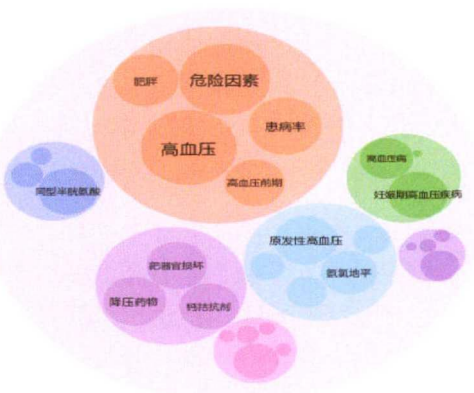


图 9 聚类中心词分布

图中每一个大圈表示一个聚类，圈内的小圈代



表聚类中的中心词,圈的大小代表聚类的规模大小或者是中心词的词频。综合分析结果:就聚类 0 来说,高血压、危险因素、患病率等词出现频繁,相关内容的研究文献数量丰富。结合其平均发文年限可知,早在 2008 年左右研究就达到一定的广度,结合关键词的分析,这一聚类研究持续时间相对较长,研究内容已具有一定的深度。考查平均发表年份和聚类规模,发现 2014 年的聚类 10 和 13 规模很小,推测是研究的主流方向,而且需要在相关关键词方向上进行纵向研究,今后的高血压研究者可以着眼于聚类 10 中的“动脉硬化”、“高胆固醇血症”和聚类 13 中的“心室功能”、“超声检查”等中心词所涉及的方面。另外,聚类 14 的规模很小,且发文年限是 2006 年,说明这一聚类中的“中医药疗法”、“清热剂”和“肝胆实火”等内容虽还没有很大的研究规模,也不是近年内所研究的重点。整体来说,高血压的各个研究领域相互渗透、相互影响,在高血压的研究内容上纵横交叉,但又不断有新的研究热点出现。

## 5 结论与建议

### 5.1 时间分布

在近 10 年的时间里,高血压科研成果持续稳定发展,且基本保持年发文量 2 000 篇以上的水平,其研究成果主要发表在专业性较强的医学或高血压领域的期刊上。

### 5.2 研究机构和作者合作

参与高血压研究的机构很多,既有一些科研实力突出的团队,又是多家并进的局面。机构之间虽有合作关系,但效果并不明显,这也是中心性低的重要原因。研究高血压的作者数量众多,彼此之间的合作程度较高,已经形成以孙宁玲、胡大一等为中心的最大研究团队,伴随着许多小团体。因此建议持续维持并扩大研究团队,同时加强团队间的合作研究。

### 5.3 关键词共现

高血压研究成果关键词结构紧凑,网络密度相对较高,但也存在部分关键词意义相重,如“高血压”和“高血压病”,不能很好地对关键词进行选

择。对每一年的关键词分析发现,证候、证素、相关性、中医症型、性别、应激性、队列研究等关键词的相关研究目前还处在起步阶段,预测是今后高血压研究的重点内容。

### 5.4 研究前沿时序

高血压研究主题较多,横向研究的广度和纵向研究的深度都已达到一定的水平,但由于其特殊性,只涉及医疗卫生领域,学科特征明显,研究前沿可以在其他学科寻找新的切入点,如借助新技术围绕“心室功能”、“超声检查”等关键词展开的研究领域可以作为进一步研究的切入点。

## 参考文献

- 1 杨功焕,孔灵芝,赵文华,等. 中国慢性病的挑战与应对 [C]. 上海: 2007 上海公共卫生国际研讨会, 2007: 28-38.
- 2 刘巍,熊兴江,王阶. 高血压病代谢紊乱及化浊平肝法的临床运用 [J]. 中国中药杂志, 2013, 38 (8): 1251-1254.
- 3 孙宁玲. 中国高血压进展与管理模式探索 [C]. 苏州. 中国心血管健康联盟一周年总结会, 2016.
- 4 Lewington S, Lacey B, Clarke R, et al. The Burden of Hypertension and Associated Risk for Cardiovascular Mortality in China [J]. JAMA Intern Med, 2016, 176 (4): 524-532.
- 5 Chen C. CiteSpace II : Detecting and Visualizing Emerging Trends and Transient Patterns in Scientific Literature [J]. Journal of the American Society for Information Science and Technology, 2009, 57 (3): 359-377.
- 6 张子石,吴涛,金义富. 基于 CiteSpace 的网络学习知识图谱分析 [J]. 中国电化教育, 2015, (8): 77-84.
- 7 成全,周兰芳. 我国微博信息聚合研究现状及路径探析-基于 CiteSpace 的可视化视角 [J]. 现代情报, 2017, 37 (3): 155-160.
- 8 王理,肖水凤,姚敏,等. 基于 CiteSpace 的医学信息学研究热点可视化分析 [J]. 中国数字医学, 2015, 10 (10): 27-29, 75.
- 9 林玲,陈福集. 基于 CiteSpace 的国内网络舆情研究知识图谱分析 [J]. 情报科学, 2017, 35 (2): 119-125.
- 10 陈兰杰,董芳. 基于知识图谱的国际竞争情报研究热点与前沿的信息可视化分析 [J]. 医学信息学杂志, 2010, 31 (8): 7-11.
- 11 范君,张艳丽. 基于 CiteSpace 的医学期刊研究文献计量分析 [J]. 淮北师范大学学报 (自然科学版), 2015, 36 (4): 88-92.