以肠道菌群为背景的结直肠癌研究可视化分析*

彭　雯　孙志广　刘丽娜　　罗　超

（南京中医药大学　南京 210029）　　（江苏省第二中医院　南京 210017）

〔**摘要**〕　采用文献分析软件 CiteSpace，对 2000－2015 年来 Web of Science 数据库收录的 731 篇以肠道菌群为背景的结直肠癌相关研究文献从年发文量、作者、地区、共引及共现等方面进行知识图谱分析。结果显示该领域研究热度呈逐年升高趋势，重点区域为美国和中国，多集中在各国高等院校中，研究热点主要是某些特定的菌群和菌群代谢产物。

〔**关键词**〕　肠道菌群；结直肠癌；CiteSpace 软件；Web of Science 数据库；可视化分析

〔**中图分类号**〕R－056　　〔**文献标识码**〕A　　〔**DOI**〕10.3969/j.issn.1673－6036.2016.12.015

Visual Analysis of Colorectal Cancer Research Based on Gut Microbiota

PENG Wen, SUN Zhi－guang, LIU Li－na; Nanjing University of Chinese Medicine, Nanjing 210029, China; LUO Chao, The Second TCM Hospital of Jiangsu Province, Nanjing 210017, China

〔**Abstract**〕　With the literature analysis software CiteSpace, the paper carries out knowledge map analysis on 731 relevant colorectal cancer research literatures based on gut microbiota included in the Web of Science database from 2000－2015 in the aspects of annual number of published papers, authors, regions, co－citation and co－occurrence. Through the result, heat degree of this field increases year by year, especially in America and China and institutions of higher learning in many countries. The research mainly focuses on some specific microbiota and microbiota metabolic products.

〔**Keywords**〕　Gut microbiota; Colorectal cancer; CiteSpace software; Web of Science database; Visual analysis

1　引言

结直肠癌是发生在人体下消化道结肠或直肠部位的恶性肿瘤，在世界范围内其发病率和死亡率均居各种癌症的前 3 位，对人体健康构成巨大的威胁。自 1975 年研究者发现无菌动物的肠癌发生率明显低于含有正常肠道菌的动物[1]以来，肠道菌群与结直肠癌发生发展的关系备受关注，基于肠道菌群的结直肠癌研究层出不穷。而分析基于肠道菌群的结直肠癌研究现状与热点，可以预测该领域的研究前景并可为今后该领域的研究方向提供参考。Web of Science 是美国汤姆森科技信息集团（Thomson Scientific）基于 Web 开发的产品，是大型综合性、多学科、核心期刊引文索引数据库，收录全球 9 000 多种久负盛名的研究期刊。CiteSpace 软

〔**修回日期**〕　2016－05－25

〔**作者简介**〕　彭雯，硕士研究生，发表论文 1 篇；通讯作者：孙志广。

〔**基金项目**〕　国家自然科学基金面上项目“大黄潜在致癌风险机制及规避策略的研究”（项目编号：81673795）；江苏省自然科学基金面上项目“基于肠道菌群探讨大黄致癌风险的机制及规避策略”（项目编号：BK20151567）。

件[2-3]是美国德雷克赛尔大学陈超美博士基于引文分析理论开发的软件，该软件被全球众多学者应用进行引文分析[4-8]。本文借助上述二者，对基于肠道菌群的结直肠癌相关研究文献进行可视化分析。

2　数据来源及方法

本研究数据来源于 Web of Science，时间跨度为 2000－2015 年，数据库选择 SCI－Expanded、SSCI、CCR－Expanded 和 IC，文献类型选择 Article。由于不同文献对“肠道菌群”及“结直肠癌”的表述存在差异，因此，本研究收集应用相对较多的表述，将检索式确定为：TS = colorectal cancer AND TS = (gut flora OR gut microbiota OR intestinal flora OR intestinal microflora OR bacteria OR microbe)，检索结果 731 条。导出每篇文献的作者、标题、来源出版物、摘要和引用的参考文献，保存成 txt 纯文本格式。基于 Web of Science 的文献分析功能对每年发文量趋势、重点机构进行统计，利用 CiteSpace 软件（版本 3.8.R1）对国家、机构、作者、关键词等进行共现分析，对文献进行共引分析。

3　分析结果

3.1　年度发文量趋势

通过对基于肠道菌群的结直肠癌研究年度发文量进行统计分析，可揭示该领域 2000－2015 年的研究文献量变化，进而可从总体上把握国际上该领域的研究趋势。结果显示基于肠道菌群的结直肠癌研究发文量总体呈逐年上升趋势。2000－2009 年，发文量上升较慢，从 2010 年开始发文量呈稳步上升趋势，到 2015 年该领域研究有了突飞猛进的发展，发文量为 16 年来最高值，见图 1。总体看来，基于肠道菌群的结直肠癌研究在国际上得到了越来越多的重视。

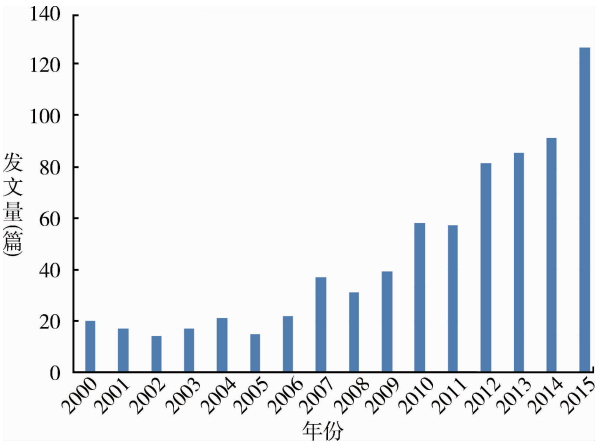

图 1　年度发文量统计

3.2　地域和机构分布

3.2.1　地域　对发表文献的地域和机构进行统计分析，可发现基于肠道菌群的结直肠癌研究领域的热点地区及重要研究基地，为发现潜在的合作国家及机构提供指引。操作 CiteSpace 软件，选择节点为 country，切片宽度为 slice = 1，阈值为 top N = 50，得到基于肠道菌群的结直肠癌研究的国家分布图谱，共有 52 个节点，46 条连线，通过调整图谱，显示发文量排名前 10 的国家见图 2，图中圆圈代表国家，圆圈直径越大，发文量越多，含有紫色圈代表与其他国家在该领域的合作较多，连线表示这些国家研究之间的关系。从图 2 可见，美国、中国、日本、德国是发文量较多的国家，美国、英国、荷兰、澳大利亚与其他国家的合作较多，其中美国的研究与其他国家关系最密切。我国的发文量虽然居第 2 位，但与发文量第 1 位的美国相比差别较大，说明我国在该领域的研究与美国相比存在一定差距。

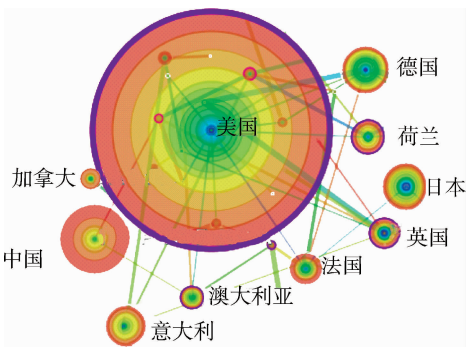

图 2　发文量国家分布

就图谱的中心性而言，在整个网络中有两个核心的中心节点，代表美国的节点中心性最大为 0.43，其次是比利时、英国，中心性分别为 0.19 和 0.17。中心性是社会网络中代表重要性的概念，某个国家的中心性越高，说明这个国家的研究越重要，起着连接其他国家研究的作用。由此可知，美国、比利时与英国在基于肠道菌群的结直肠癌研究中处于非常重要的地位。从发文量的突增性（bursts）来看，我国为 7.13，排名第 1，德国为 4.59，排名第 2。发文量突增性是反映发文量增长的指标，突增性越大，说明集中性的研究越多。我国的发文量突增发生在 2013－2015 年，而德国的发文量突增发生在 2006－2007 年，说明我国是近年来基于肠道菌群的结直肠癌研究增长最快的国家。

3.2.2　机构　利用 Web of Science 的统计分析功能对基于肠道菌群的结直肠癌研究机构进行统计，展示发文量前 10 的机构，见图 3。可以看出该领域的研究主要集中在各国的高等院校中，在排名前 10 的机构中有 9 家为大学。发文量最多为美国哈佛大学共 21 篇，其次是美国北卡罗来纳大学共 20 篇，我国的上海交通大学发文量居第 3 位为 17 篇。

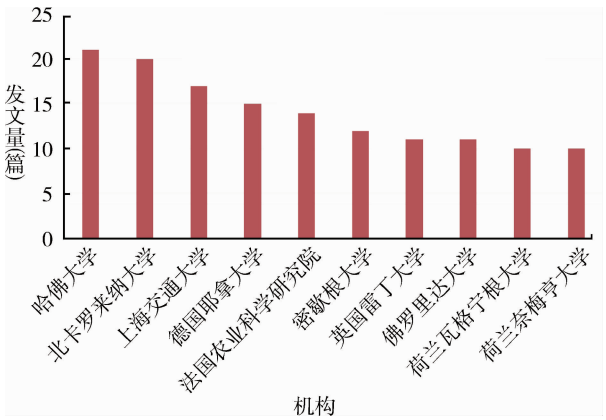

图 3　发文量前 10 的机构

3.3　作者分析

基于 Web of Science 的数据分析功能，统计本次检索数据中发文量前 10 的作者及其所在单位。了解某一领域的作者分布，能够发现该领域的顶尖研究者，可从他们的研究内容了解当前研究的深度和广度，预测科学研究的前沿，指引未来研究方向。表 1 中列出了发文量排名前 10 位的作者，这些作者来自美国的居多，有 7 位，荷兰学者 1 位，德国学者 2 位。Boleij Annemarie 教授与 Tjalsma Harold 教授是重要的合作伙伴，两人曾合作发表了该领域具有代表性的文献，如著名的“driver－passenger 模型”，其认为某些肠道固有菌（driver）引起了肠上皮的 DNA 损伤，是癌变过程的始动因素，癌变过程中肠道微环境发生变化，这种微环境有利于另外一些菌（passenger）的定殖，这一模型假说被用来解释肿瘤组织中益生菌风度较高而潜在致病菌却减少的现象[9]。

表 1　发文量前 10 作者的机构与国家

作者	发文量（篇）	机构与国家
Boleij Annemarie	10	约翰霍普金斯大学医学院医学系，美国
Tjalsma Harold	10	奈梅亨大学医学中心，荷兰
Glei Michael	9	耶拿大学营养学院营养毒理系，德国
Jobin Christian	9	北卡罗来纳大学药理学系，美国
Pool－Zobel BL	8	耶拿大学营养学院营养毒理系，德国
Abreu Maria T	7	迈阿密大学米勒医学院医学系，美国
Fox James G	7	麻省理工学院比较医学部，美国
Gaskins H. Rex	7	伊利诺斯大学营养科学系，美国
Huycke Mark M	7	退伍军人事务部医疗中心，美国
Kato Ikuko	7	韦恩州立大学医学院 Karmanos 肿瘤研究所，美国

3.4　关键词共词网络分析

3.4.1　共现分析　关键词是一篇文章的核心要素，通过关键词共词分析可以综合分析所确定的时间段内文献中的热点研究词汇，从而推断出此段时期该领域的研究热点。出现频次较高的术语，是对这些关键词的补充。本研究通过分析 2000－2015 年

基于肠道菌群的结直肠癌研究文献的关键词及术语，来推测该领域 16 年中最值得关注的热点。运行 CiteSpace 软件进行共现分析，选择切片为 slice = 1，术语类型为 nounphrases，网络节点类型为 keyword、term，阈值 top$N$ = 30，生成由关键词和术语组成的关键词共词网络图谱，共有 194 个节点，234 条连接线。对表达形式不同但意义完全相同的词汇进行合并，如"colorectal - cancer"和"colorectal cancer"、"colon - cancer"和"colon cancer"等，最后得到图 4。图中字号越大，表示该词的共现频次越高。可以看出，除去与检索策略相关的"colorectal cancer"和"gut microbiota"，共现频次最高的是"inflammatory bowel disease"，即炎症性肠病。与炎症性肠病密切相关的溃疡性结肠炎、克罗恩病、肠易激综合征、肠炎等词汇共现频次也较高，可见在基于肠道菌群的结直肠癌研究中，各种炎症性肠病等是研究的热点。

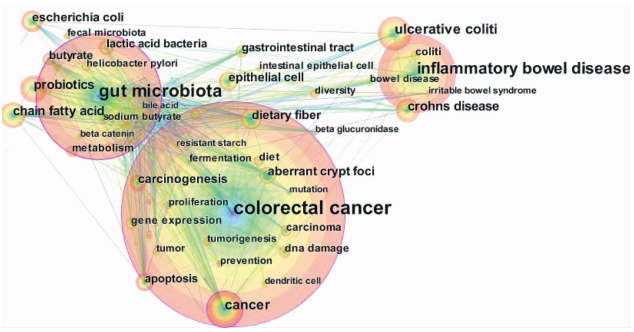

图 4　关键词及术语共现分析

据文献报道患有炎症性肠病的人群，其肠癌发病率是健康人群的 2 ~ 4 倍[10]。从目前的分析可以推测，肠道菌群可能通过影响炎症性肠病的病情发展而对结直肠癌产生影响。其次，可以表现大肠杆菌、乳酸菌、幽门螺旋杆菌、益生菌、丁酸盐、链脂肪酸、代谢等词汇出现，这些细菌和细菌代谢产物等在基于肠道菌群的结直肠癌研究扮演了举足轻重的角色，某些特定的细菌或细菌代谢产物可能参与了结直肠癌的演进过程。另外，与异常隐窝灶、凋亡、树突状细胞、DNA 损伤、基因表达、突变、肿瘤、癌症等词汇的共现频次也很高，这些术语主要与描述结肠癌癌变的机制有关。

3.4.2　突显分析　对关键词进行突显分析，能够发现某一年份突增的研究热点，并且可以看出该热点持续出现的年份。对查找到的 731 篇文献的关键词进行突显分析，合并类似的词，如"gut microbiota"、"gut microbiome"等，得出 10 个突显词见图 5，可见膳食纤维（dietary fiber）和短链脂肪酸（short - chain fatty acids）都是 2000 年突增的研究热点，持续到 2007 和 2008 年，是持续研究时间最长的热点词汇。在这一时期基于肠道菌群的结直肠癌研究多偏向于研究饮食中的膳食纤维和短链脂肪酸在这其中发挥的作用。溃疡性结肠炎（ulcerative colitis）和肠道炎症（intestinal inflammation）作为突显词出现，说明溃疡性结肠炎在内的炎症性肠病在该领域中曾一度成为研究的热点，反复的炎症刺激，肠黏膜的过度修复及瘢痕形成，极易导致肠癌的发生。而近几年菌群紊乱（dysbiosis）、梭杆菌（fusobacterium）、粪便菌群（fecal microbiota）成为新的研究热点。肠道菌群的紊乱，某些特殊菌群的研究（如梭杆菌）可能将继续成为该领域研究的前沿。是否能找出类似于促进胃癌发生发展的幽门螺旋杆菌的细菌来解释结直肠癌的发展进程，将是科研工作者下一步的探索方向。

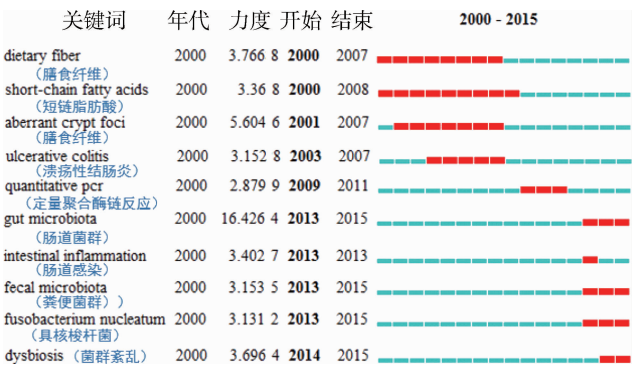

图 5　关键词突显分析

3.5　文献共引分析

3.5.1　重点文献分析　设置 CiteSpace 软件，选择切片 slice = 1，节点为 citedreference，阈值 top$N$ = 30，得到基于肠道菌群的结直肠癌研究的文献共引网络图谱，共有 360 个节点，479 条连线，图略。引用频次前 10 的文章，见表 2。

表 2　引用频次前 10 的文章

引用频次	文献索引
51	Kostic AD, 2012, GENOME RES, V22, P292, DOI 10.1101/gr.126573.111
49	Wu SG, 2009, NAT MED, V15, P1016, DOI 10.1038/nm.2015
46	Sobhani I, 2011, PLOS ONE, V6, DOI 10.1371/journal.pone.0016393
46	Arthur JC, 2012, SCIENCE, V338, P120, DOI 10.1126/science.1224820
45	Castellarin M, 2012, GENOME RES, V22, P299, DOI 10.1101/gr.126516.111
35	Marchesi JR, 2011, PLOS ONE, V6, DOI 10.1371/journal.pone.0020447
35	Shen Xiang Jun, 2010, GUT MICROBES, V1, P138, DOI 10.4161/gmic.1.3.12360
33	Eckburg PB, 2005, SCIENCE, V308, P1635, DOI 10.1126/science.1110591
30	Wang TT, 2012, ISME J, V6, P320, DOI 10.1038/ismej.2011.109
29	Uronis JM, 2009, PLOS ONE, V4, DOI 10.1371/journal.pone.0006026

从被引用频次来看，2000－2015 年基于肠道菌群的结直肠癌研究在国际上被引次数最高的文献是 Aleksandar D. Kostic 于 2012 年发表在《基因组研究》（*Genome Research*）的文章，题为“运用基因组方法分析梭杆菌与结肠癌之间的相关性”，其引用频次为 51 次。文章作者应用基因组学方法进行了多项试验，深入探究了梭杆菌种属的变化对结直肠癌发生发展的影响。研究引用轨迹发现，该文从 2012 年起处于持续被引用中，2014、2015 年分别被引用达 16 次之多。能够看出基于梭杆菌的结直肠癌研究受到了众多学者的推崇。其次，被引用较多的还有 Shaoguang Wu 发表的“人类结肠共生菌通过 TH17 反应的激活促进结肠肿瘤的发展”，载于《自然－医学》（*Nature Medicine*）及 Janelle C. Arthur 发表的“肠道炎症作用于癌症——包括菌群的作用”，载于《科学》，被引频次分别为 49 和 46 次，这两篇文章分别从 Th17 细胞和炎症方面介绍了肠道菌群对结直肠癌发生发展的作用[11－12]。

3.5.2　文献分类　对共引文献进行聚类分析，得到 Modularity Q = 0.826 7，Mean silhoutte 值为 0.710 7，Modularity 表示网络的模块度，值越大表示网络的聚类结果越好。Silhouette 值是用来衡量网络同质性的指标，越接近 1 反映网络的同质性越高。采用 timeline 视图，基于肠道菌群的结直肠癌研究的研究主要分为 8 个大簇类，见图 6。

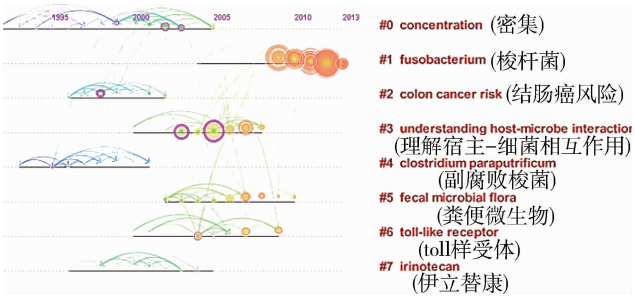

图 6　聚类分析（Timeine 视图）

在图 6 中，每条实线代表的是某一簇类所在的时间跨度，节点文章圆圈厚度越大，引用次数越多，外圈为紫色的节点是连接各簇类之间的关键性文章。总体而言，这 8 类中#0 类密集（concentration）时间跨度较大；#3 类理解宿主－细菌相互作用（understanding host－microbe interaction）所涉及的交叉性节点文章较多，可见理解宿主与菌群之间的相互作用对于该领域的研究具有重要意义；引文较集中分布的类别为#1 梭杆菌（fusobacrerium），说明梭杆菌仅为今年的研究热点；从图中还可以发现，toll 样受体（toll－like receptor）和伊立替康（irinotecan）也成为主要类别，可见在发病机制研究中，模式识别受体中的 toll 样受体受到关注；在治疗方面伊立替康等肠癌治疗药物也一再引起研究者的兴趣。这里对近几年研究较集中的#1"fusobacrerium" 簇类中具有代表性的文献“在结直肠癌肿瘤微环境中，微生物毒力基因起着重要作用”进行进一步分析。该文献来自美国明尼苏达大学的学者 Michael B. Burns，于 2015 年发表在《基因组医学》（*Genome Medicine*）杂志上。该文章利用相关试验，研究毒性基因在结直肠癌肿瘤微环境中的作用。研究中采用基因组测序技术对 44 例原发结直肠癌组织和 44 例匹配的正常结肠组织中微生物进行分析，

以确定与肿瘤相关的微生物的富集程度。结果显示肿瘤微环境中，微生物多样性有所增加，尤其是梭杆菌和普罗威登斯菌的风度变化。预测毒力相关基因的富集，可能取决于梭杆菌和普罗维登西亚的基因组[13]。

4 结语

本文以 2000 – 2015 年 Web of Science 收录的基于肠道菌群的结直肠癌研究相关文献（共 731 篇）为数据来源，借助文献统计分析与 CiteSpace 知识图谱的方法，对国际上基于肠道菌群的结直肠癌研究现状与热点进行了分析。从分析结果来看，基于肠道菌群的结直肠癌研究热度呈逐年增长趋势，美国和中国研究成果丰硕，该领域的主要研究机构集中在各国的大学中，顶尖学者主要来自美国；而在研究热点方面，主要集中在某些特定的菌群（梭杆菌等）和菌群代谢产物（如丁酸）上，同时炎症及免疫相关机制的研究也持续被众多学者关注。可以预测在未来的一段时间中，基于肠道菌群的结直肠癌研究仍然会继续保持研究热度肠道菌群与结直肠癌发生发展的关系也会逐渐明朗，科学真理将会被进一步发现。

参考文献

1 Vannucci L, Stepankova R, Kozakova H, et al. Colorectal Carcinogenesis in Germ – free and Conventionally Reared Rats: different intestinal environments affect the systemic immunity [J]. Int J Oncol, 2008, 32 (3): 609 – 617.

2 陈悦，胡志刚. 引文空间分析原理与应用：CiteSpace 实用指南 [M]. 北京：科学出版社，2014.

3 陈超美，侯剑华. CiteSpaceII：科学文献中新趋势与新动态的识别与可视化 [J]. 情报学报，2009，28 (3)：401 – 421.

4 郝维琴，吴冬梅，周维军. 国内高血压病研究的知识图谱分析 [J]. 实用医药杂志，2015，32 (5)：468 – 470.

5 许吉，方肇勤. 基于 SCI 数据的原发性肝癌可视化分析 [J]. 数理医药学杂志，2015，28 (6)：853 – 856.

6 张惠荣，何冰，刘兴太. 基于文献共引网络的救援医学研究可视化分析 [J]. 图书馆工作与研究，2012，(11)：99 – 101.

7 郭妍，周维军，程保青. 老年糖尿病文献可视化分析 [J]. 实用医药杂志，2015，32 (4)：369 – 370.

8 李嵬. 鼻咽癌研究的科学知识图谱分析〔D〕. 长沙：中南大学，2014：36 – 38.

9 Tjalsma H, Boleij A, Marchesi JR, et al. A Bacterial Driver – passenger Model for Colorectal Cancer: beyond the usual suspects [J]. Nat Rev Microbiol, 2012, 10 (8): 575 – 582.

10 Farraye FA, Odze RD, Eaden J, et al. AGA Medical Position Statement on the Diagnosis and Management of Colorectal Neoplasia in Inflammatory Bowel Disease [J]. Gastroenterology, 2010, 138 (2): 738 – 745.

11 Wu S, Rhee KJ, Albesiano E, et al. A Human Colonic Commensal Promotes Colon Tumorigenesis Via Activation of T Helper Type 17 T Cell Responses [J]. Nat Med, 2009, 15 (9): 1016 – 1022.

12 Arthur JC, Perez – Chanona E, Muhlbauer M, et al. Intestinal Inflammation Targets Cancer – Inducing Activity of the Microbiota [J]. Science, 2012, 338 (6103): 120 – 123.

13 Burns MB, Lynch J, Starr TK, et al. Virulence Genes are a Signature of the Microbiome in the Colorectal Tumor Microenvironment [J]. Genome Med, 2015, 7 (1): 55.

欢迎订阅 欢迎赐稿