

基于期刊与专利文献计量分析的 慢性粒细胞白血病热点判断

曲红¹ 张英杰²

(1. 广州番禺区中心医院, 广东广州 510530; 2. 中国科学技术信息研究所, 北京 100038)

摘要: 慢性粒细胞白血病是一种获得性造血干细胞恶性克隆性疾病, 临床表现以贫血、反复感染、出血为主, 对患者的生命安全构成了严重的威胁。文章基于1955—2017年慢性粒细胞白血病专利文献数据和1973—2017年慢性粒细胞白血病期刊文献数据, 运用了文献分析方法和知识图谱方法, 从期刊和专利文献增长数量、机构增长数量、技术主题等进行了多角度研究。研究发现, 目前慢性粒细胞白血病期刊文献发表数量呈现明显的增长趋势, 技术主题方向处于成熟期; 从研究热点来看, 期刊和专利表现出截然不同的方向。从两类文献揭示的热点主题来看, 期刊在热点主题的揭示方面, 更接近临床应用, 而专利更侧重于各类并发症、各类相关肿瘤治疗方案。

关键词: 慢性粒细胞白血病; 期刊; 专利分析; 文献分析; 热点分析

中图分类号: R733.7; G306

文献标识码: A

DOI: 10.3772/j.issn.1674-1544.2018.05.014

Analysis of the Application Based on the Journal Literature and Patent by Example of Hot Topic Determination of Chronic Myeloid Leukemia

QU Hong¹, ZHANG Yingjie²

(1. Guangzhou Panyu District Central Hospital, Guangzhou 510530; 2. Institute of Science and Technology Information of China, Beijing 100038)

Abstract: Chronic myeloid leukemia is a malignant clonal disease of acquired hematopoietic stem cells. The clinical manifestations are mainly anemia, repeated infection, and hemorrhage, which pose a serious threat to the patient's life safety. Based on the 1955-2017 chronic granulocyte leukemia patent literature data and the 1973-2017 chronic granulocytic leukemia journal literature data, this paper uses literature analysis methods and knowledge graph analysis methods, from the number of patent documents, the number of institutional growth, technical topics, etc. A multi-angle study. The main conclusion is that the number of published literatures of chronic myeloid leukemia journals has shown a significant growth trend, and the technical subject direction is at maturity. From the research hotspots, journals and patents show very different directions. From the hot topics revealed by the two types of literature, journals are closer to clinical applications in the disclosure of hot topics. While patents focus on various complications and various types of diseases, which is related tumor treatment options.

Keywords: chronic myeloid leukemia, journal, patent analysis, literature analysis, hot topic analysis, knowledge graph analysis

作者简介: 曲红 (1980—), 女, 广州番禺区中心医院主治医师, 研究方向: 血液病学; 张英杰 (1979—), 男, 中国科学技术信息研究所副研究员, 研究方向: 科技信息资源管理、科技大数据服务。

资助项目: 国家重点研发计划专项项目“知识产权信息智能采集及深加工技术研究与应用示范”(2017YFB1401902); 中国科学技术信息研究所重点工作“面向中信所资源大数据建设的多源异构数据库内容获取与融合平台建设”(ZD2017-09)。

收稿时间: 2018年8月27日。

科技期刊作为最新科技成果的传播媒体,是科技人员发表论文、开展学术交流的重要通道,在很多情况下,科技期刊是科研人员科技成果的首发选择。据统计,我国自然科学类科技期刊约占全部期刊的49.5%,这些期刊极大地促进了各类新科技的传播、共享^[1]。

据WIPO统计,专利信息包含了世界科技信息的90%~95%,在科学研究中利用专利信息不仅可以减少60%的研发时间和40%的研发费用^[2],而且相对于其他文献形式,专利更具有新颖、实用的特征。世界各国对技术的保护意识越来越强烈,纷纷对已有技术成果申请专利,进行关键技术保护。

近年来,随着医学科技水平的大幅提高,人们已经研发出各种治疗慢性粒细胞白血病(Chronic Myeloid Leukemia, CML)的药物,慢性粒细胞白血病已从不治之症转变为有正常生存期的疾病,治疗目标已从延长生存期转向维持生活质量。传统化疗药物见效快,研究者正在把目光集中在靶向BCR-ABL融合蛋白上,酪氨酸激酶抑制剂针对性强,副作用小,被越来越多的人所青睐。为此,本文围绕慢性粒细胞白血病的热点研判,分别选择该主题的中文期刊数据和德温特专利数据,在探究其发展状况和趋势的基础上,深入地对慢性粒细胞白血病技术进行挖掘,试图通过分析期刊、专利两类数据在热点发现揭示方面的差异,找出慢性粒细胞白血病未来的重点发展方向。

1 数据来源与研究方法

1.1 数据集构建

本文选择“慢性粒细胞白血病”or“Chronic Myeloid Leukemia”作为检索词,分别抽取了两部分数据形成分析用数据集,其中的专利数据来源于德温特专利数据库(Derwent Innovations Index),期刊数据来源于中外文科技文献元数据仓储知识库。上述数据经过检索清洗后,形成论文数据5514条(1973—2017年),专利数据539条(1955—2017年)。

德温特专利数据库支持快速而精确的专利和

引文检索,内容涵盖化学、电气、电子和机械工程等领域。借助附加的描述信息和编码以及可追溯到1963年的专利收录内容,使用者能够快速了解某一专利的重要性及其与其他专利的关系。

中外文科技文献元数据仓储知识库的建设,得到了国家863计划的资助和支持,覆盖10种资源类型,涵盖260余个中外文主流数据厂商数据库,累计收录7.3亿条科技文献元数据^[3]。数据仓储知识库为海量科技文献资源整合及学术搜索提供了可能,基于中外文元数据仓储知识库,实现了学术搜索与全文资源调度、作者科研协作关系分析、机构科研协作关系分析、学科主题知识脉络与学术发展趋势分析。

1.2 研究方法

目前,围绕研究热点发现和研判,国内外学者利用的数据类型包含有期刊、专利、网络文献、科技规划、基金项目等。用于开展热点发现的方法有基于文献引用的文献计量法,采用Histcite软件构建引用关系研究数据,在对热点文献进行解读的基础上形成对研究热点的研判;基于关键词/主题词的共词分析法,基本原理是对文献中的关键词进行共现统计,进而利用不同的聚类算法,分析关键词所代表的学科和主题变化;基于数据挖掘的知识图谱法,比较典型的是采用CiteSpace等软件,结合Web of Science数据库,形成相关领域知识图谱,揭示热点主题演变。

在具体实践方面,孟凡赛^[4]使用有序聚类方法,利用期刊主题词集合,进行学科领域主题演化研究;刘彤^[5]从专利的角度对电力产业技术领域进行分析,揭示该领域的技术热点和区域竞争格局;周群^[6]等将最新论文文献与基于科技媒体的新闻报道、评论和观点文章进行结合,提出一种基于科技媒体视角的研究前沿识别与探测方法,有效揭示研究前沿的重大研究突破;司林波等^[7]运用CiteSpace文献计量可视化软件,选择8种公共管理研究期,构建文献关键词的知识图谱,探测到近十年国内公共管理研究的热点;李侗桐等^[8]也对网络文献和期刊文献在揭示研究热点方面的差异进行了比对分析。

围绕“慢性粒细胞白血病”主题,景晓宇

等^[9]以CNKI论文数据为基础,采用共词分析的方法,对我国近年来慢性粒细胞白血病的研究进展进行探讨。仓顺东等^[10]依据第50届美国血液病学年会(ASH)的会议论文,对伊马替尼(Imatinib)的出现及其对慢性粒细胞白血病(CML)患者的治疗效果进行了探讨。经过文献梳理,发现围绕该主题领域的热点探讨更多的是采用综述或案例研究的方式,基于某种药物或治疗方案,开展最新进展或方案效果研究。如段丽芳等^[11]通过分析浆细胞样树突状细胞这一免疫佐剂,分析了其与慢性粒细胞白血病免疫治疗进展。

本文拟在医学领域常规案例研究的基础上,结合科技期刊和专利的科技创新传播特征,通过大量的慢性粒细胞白血病相关文献的分析充分了解慢性粒细胞白血病发展状况。在此基础上,采用知识图谱的方式,对专利和期刊文献耦合,形成对该主题的热点研判,进而比较期刊和专利两种文献类型在揭示热点主题方面的应用差异。

2 研究趋势与主要研究参与者

2.1 研究趋势

对慢性粒细胞白血病1955—2017年期刊文献与专利文献年增长趋势的对比进行分析,结果如图1所示。从总体上看,慢性粒细胞白血病期刊文献和专利文献数量都呈上升趋势,但期刊文献年增长速度较快,专利文献年增长速度较

慢。其中,1995—2000年,期刊文献与专利文献年增长数量均较少,这说明慢性粒细胞白血病相关技术研究水平较低,处于技术发展起步期与瓶颈期。2001—2014年期刊文献和专利文献数量均出现了较快的增长,且增长速度较快,在一定程度上可以说明这一时间段技术发展较为迅速。2015年之后期刊文献和专利文献年增长数量有所下降,说明慢性粒细胞白血病相关基础研究趋于成熟和完善,基础技术不断提高。

图2、图3分别对慢性粒细胞白血病1955—2017年期刊论文作者与专利发明人年增长数量及论文作者单位与专利权人机构年增长数量进行对比分析,结果发现,同期期刊文献与专利文献年增长数量对比的结果基本一致,但专利发明人在2002—2008年快速增长,与同期期刊作者数量反而缓慢下降表现出截然不同的特征。

2.2 主要研究参与者

2.2.1 主要研究人员

通过对期刊、专利进行统计,分别取期刊论文发表量超过4篇的作者和专利申请量超过2件的发明人,如表1所示,可以发现,期刊文献作者排在前面的有林伟、姚尔固、李百泉、薛永泉、彭光斌、杨清明、孟凡义、宁辉等,专利文献中的专利发明人排在前面的有KEITH C、DEMIN P、ROUNOVA O、GRUNBERGER T、GAUDINO J J、FOLEY M A、LEE M S等。可以发现,在慢性粒细胞白血病领域中,期刊文献作者与专利文



图1 期刊文献与专利文献年增长数量对比图

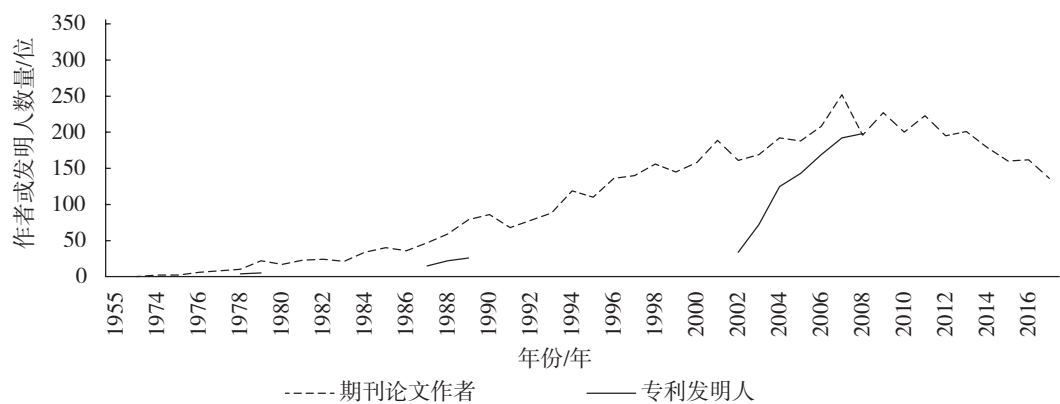


图2 期刊作者与专利发明人年增长数量对比图

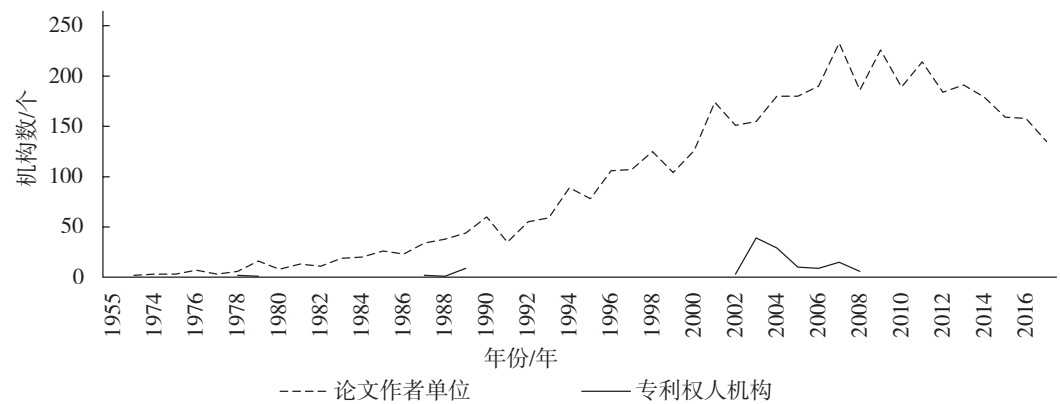


图3 论文作者单位与专利权人机构年增长数量对比图

表1 期刊文献作者与专利发明人对应表

编号	期刊文献作者	数量/篇	专利发明人	数量/件
1	林 伟	6	KEITH C	4
2	姚尔固	6	DEMIN P	3
3	李柏泉	6	ROUNOVA O	3
4	薛永权	5	GRUNBERGER T	3
5	彭光斌	5	GAUDINO J J	3
6	杨清明	5	FOLEY M A	3
7	孟凡义	5	LEE M S	3
8	宁 辉	5	OHNO T	3
9	谢 波	4	LI Y	3
10	尤育初	4	MOUSSY A	3

献中的专利发明人并没有对应,说明发表期刊论文的作者并没有转化为专利的发明人。

2.2.2 主要研究机构

通过统计,分别取期刊论文发表量超过4篇的作者单位和专利申请量超过2件的专利权人,如表2所示。可以发现,期刊所属机构排在前面的有中国医学科学院、中国协和医科大学血液学研究所、血液病医院,沈阳军区总医院血液

科,同济医科大学附属同济医院,解放军总医院血液科等机构,专利所属机构排在前面的有ICOS CORP(ELIL-C)、COMBINATORX INC(Comb-Non-standard)、OHNO T(OHNO-Individual)、LI Y(LIYY-Individual)、AB SCI(ABSC-Non-standard)、KEITH C(KEIT-Individual)、HOSPITAL FOR SICK CHILDREN(HOSP-Non-standard)、JANSSEN PHARM NV(JANC-C)等机构。与期刊文献作者与专利发明人对应表结构相似,在表2中,论文作者所属机构与专利的专利权人机构也没有相同的机构对应,在慢性粒细胞白血病领域中,发表的专利并不属于期刊所属的机构。同时,通过分析相关专利表现,主要专利还是在发达国家中。与发达国家相比,我国在慢性粒细胞白血病领域的研究中还存在一定差距。

3 研究热点对比分析

慢粒起病缓慢,病程长,传统治疗不能改

表 2 论文作者所属机构与专利权人机构对应表

编号	论文作者单位	数量/篇	专利权人机构	数量/件
1	中国医学科学院	27	ICOS CORP (ELIL-C)	4
2	沈阳军区总医院	9	COMBINATORX INC (COMB-Non-standard)	4
3	同济医科大学附属同济医院	8	OHNO T (OHNO-Individual)	3
4	解放军总医院	8	LI Y (LIYY-Individual)	3
5	北京医科大学人民医院	8	AB SCI (ABSC-Non-standard)	3
6	南方医科大学南方医院	8	KEITH C (KEIT-Individual)	3
7	军事医学科学院附属医院	7	HOSPITAL FOR SICK CHILDREN (HOSP-Non-standard)	2
8	苏州医学院	7	JANSSEN PHARM NV (JANC-C)	2
9	广东省人民医院	6	JANSSEN PHARM INC (JANC-C)	2
10	上海第二医科大学附属瑞金医院 上海血液学研究所	6	LI C J (LICJ-Individual)	2

变大多数患者存活期。异基因造血干细胞移植 (Allo-SCT) 则一般考虑应用于靶向药物治疗失败者。涉及的基础药物有羟基脲^[12]、白消安 (马利兰)^[13]、甲异靛^[13]、格列卫 (伊马替尼)^[14]；代表方案有化疗和基因干扰素工程治疗。联合用药是慢粒治疗的趋势。合理的组合能提高疗效，减少副作用，延长生存期。为此，本文通过对期刊论文、专利文本进行主题抽取，形成了若干热点主题，表 3 列出了慢性粒细胞性白血病的主要研究高频关键词。图 4 和图 5 分别是慢性粒细胞性白血病期刊和专利的热力图。热力图显示了离散数据 (事件或事物) 的分布及其相互关系，通常以一张具有显著颜色差异图片的方式呈现最终结果，颜色的深浅代表了该主题的发生频率或分布的密度大小^[15]。

通过比较两者的高频关键词可以看出，期刊热点词主要集中在白血病、粒细胞、细胞、基因、患者、骨髓、干细胞、疗效、急性等方面。

表 3 慢性粒细胞性白血病的主要研究热点关键词统计

期刊文献关键词	数量/次	专利文献关键词	数量/次
白血病	3996	1-6c alkyl	53
粒细胞	3166	multiple myeloma	29
细胞	784	chronic lymphocytic leukemia	28
基因	653	acute myelocytic leukemia	26
患者	512	pancreatic cancer	26
骨髓	382	chronic myelocytic leukemia	26
干细胞	339	breast cancer	25
疗效	335	prostate cancer	25
急性	317	renal cell carcinoma	24

专利热点词主要集中在 1-6C 烷基、多发性骨髓瘤、慢性淋巴细胞白血病、急性髓细胞性白血病、胰腺癌、慢性粒细胞白血病、乳腺癌、前列腺癌、急性淋巴细胞白血病等方面。而期刊、专

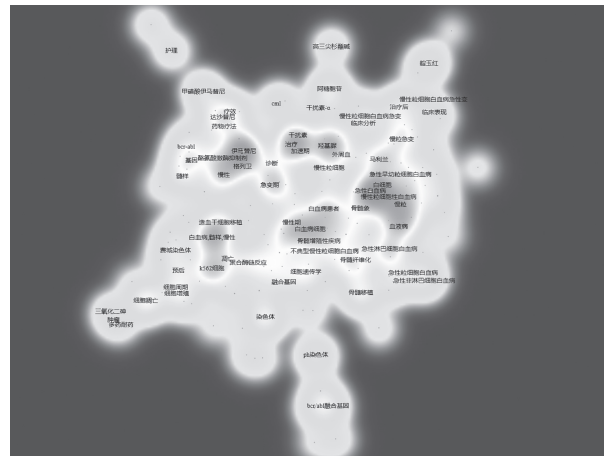


图 4 期刊论文热力图

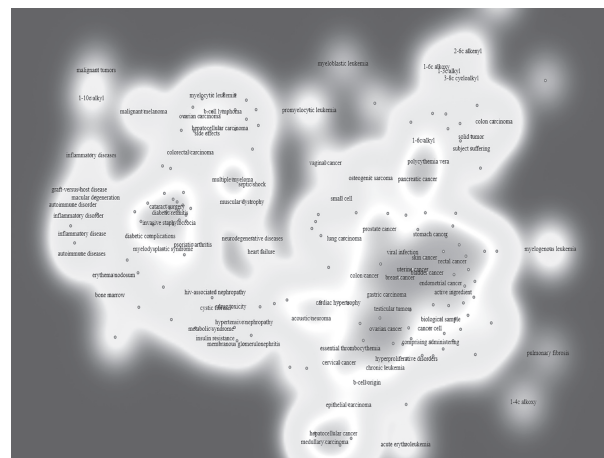


图 5 专利文献热力图

刊两种文献相同的热点词主要集中在慢性粒细胞白血病、急性、骨髓等主题上。

通过高频词与两者的热力图进行对比发现,期刊热力图大致可以划分为3个热点区域:一是围绕慢性期,主要聚焦慢性粒细胞性白血病主题的细胞遗传学领域,开展了基于不同型慢性粒细胞性白血病的治疗,涉及骨髓增殖性疾病、血液病等不同概念,也涉及马里兰等药物;二是围绕加速期和急变期,涉及干扰素、络氨酸激酶抑制剂,突出了伊马替尼、达沙替尼等药物;三是涉及造血干细胞移植主题,有聚合酶链式反应、三氧化二砷等相关主题。而专利热力图则主要划分为两个热点区域:一是慢性粒细胞性白血病并发的病毒性感染、皮肤癌、胃癌、丸肿瘤、结肠癌等各类癌症;二是围绕糖尿病性视网膜炎、多发性骨髓瘤等。

4 结论

本文基于1955—2017年慢性粒细胞白血病专利文献数据和1973—2017年慢性粒细胞白血病期刊文献数据,运用文献分析方法和知识图谱方法,从专利文献增长数量、机构增长数量、技术热点主题等进行多角度研究。围绕慢性粒细胞白血病主题,分析专利申请数量的增长趋势发现,该主题目前基本处于成熟期;期刊发表则在2014年前后达到峰值,之后呈现缓慢下降的趋势。

总体来说,期刊和专利都具有科学性、专业性、连续性特点,但期刊在新主题传播方面更为及时,这取决于二者不同的发布机制,决定了二者在时效性的差异。期刊在慢性粒细胞性白血病这一主题传播方面,比专利文献表现得更为连续、更为及时。而从上述两类文献揭示的热点主题来看,期刊在热点主题的揭示方面,更接近临床应用,如各类治疗药物名称都出现在了期刊论文中,而专利更侧重于各类并发症、各类相关肿瘤治疗方案。

参考文献

[1] 关于准确把握科技期刊在学术评价中作用的若干意

见[EB/OL].[2015-11-11]. <http://www.gapp.gov.cn/news/1663/268504.shtml>.

- [2] 专利信息战略[EB/OL].[2018-06-21]. <https://baike.baidu.com/item/%E4%B8%93%E5%88%A9%E4%BF%A1%E6%81%AF%E6%88%98%E7%95%A5/12748338?fr=aladdin>.
- [3] 中外文科技文献元数据仓储知识库[EB/OL].[2018-06-21]. <http://www.infosoft.cc/SubPage/DisplayAbstracte653.html?R=a18de381-84eb-45ef-a51f-bd7a50dc6fe1&D=26&T=RMSCloud%E4%BA%91%E6%95%B0%E6%8D%AE>.
- [4] 孟凡赛. 基于有序聚类的期刊主题演化和热点发现研究[J]. 情报探索, 2018(5): 68-71.
- [5] 刘彤, 苏可, 蒋贵凰. 基于专利的电力技术热点领域分析[J]. 科技视界, 2014(31): 29-30.
- [6] 周群, 周秋菊, 冷伏海. 基于科技媒体视角的研究前沿识别方法研究与实证[J]. 现代情报, 2018(2): 62-67, 74. DOI: 10.3969/j.issn.1008-0821.2018.02.010.
- [7] 司林波, 李雪婷, 孟卫东. 近十年中国公共管理研究的热点领域和前沿主题: 基于八种公共管理研究期刊2006—2015年刊载文献的可视化分析[J]. 上海行政学院学报, 2017, 18(2): 104-111.
- [8] 李侗桐, 冯秋蕾, 田佳. 科技期刊研究热点与网络关注热点的对比分析[J]. 中国科技期刊研究, 2017(11): 992-996. DOI: 10.11946/cjstp.201704130259.
- [9] 景晓宇, 鲍珊, 董天, 等. 我国慢性粒细胞白血病研究现状的共词分析[J]. 临床血液学杂志, 2016(1): 51-54.
- [10] 仓顺东, 宋永平, LIU Delong. 慢性粒细胞白血病治疗新进展及第50届美国血液病学会年热点[J]. 白血病·淋巴瘤, 2009, 18(5): 304-308.
- [11] 段丽芳, 张晓芹. 浆细胞样树突状细胞与慢性粒细胞白血病免疫治疗进展[J]. 医学研究杂志, 2014(1): 12-13, 134.
- [12] 杨英姿. 中西医治疗慢性粒细胞白血病进展[J]. 实用中医药杂志, 2010, 26(6): 445-446.
- [13] 姚锡珍, 仇惠英, 夏学鸣. 甲异靛治疗慢性粒细胞白血病10例临床观察[J]. 苏州大学学报: 医学版, 1996(1): 96.
- [14] 邓冬英, 邓玮华, 叶莹莹. 甲磺酸伊马替尼治疗慢性粒细胞白血病13例护理[J]. 福建医药杂志, 2010(3): 151-152. DOI: 10.3969/j.issn.1002-2600.2010.03.095.
- [15] 刘玉琴, 汪雪锋, 雷孝平. 科研关系构建与可视化系统设计与实现[J]. 图书情报工作, 2015, 59(8): 103-110.