

生物样本库研究状况及发展探讨^{*}

高雅洁 程 芸 陈瑞安 寿晓玲 蔡国龙

(浙江医院 杭州 310030)

〔摘要〕 对近 20 年关于生物样本库 SCIE 文献进行知识图谱可视化分析,揭示生物样本库发展历程、合作网络,分析其研究热点及演化路径,为科研人员提供未来研究方向和思路。

〔关键词〕 生物样本库;知识图谱;分析

〔中图分类号〕 R-058 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2021.06.007

Discussion on the Study Status and Development of Biobank GAO Yajie, CHENG Yun, CHEN Ruian, SHOU Xiaoling, CAI Guolong, Zhejiang Hospital, Hangzhou 310030, China

〔Abstract〕 The paper conducts a visualization analysis of knowledge graph on the SCIE literature of the biobank in the past 20 years, reveals the development history and cooperation network of biobank, analyzes the study hotspots and evolution paths, and provides the future study directions and ideas for researchers.

〔Keywords〕 biobank; knowledge graph; analysis

1 引言

生物样本库又称生物银行(biobank),主要用来标准化收集、处理、储存健康的或疾病生物体的

生物分子、细胞、组织和器官等样本^[1]。其建设初衷是整合有限生物样本库资源,实现将生物样本资源转化为改善医疗健康所需数据资源,促进多学科与跨专业协同合作^[2]。随着转化医学和精准医学发展,生物样本库已成为众多科研成果应用到临床、实现转化医学的重要保证^[3],也是驱动精准医学、改善医疗服务的基本工具。本文通过文献计量法、共现分析法、词频统计法和知识图谱可视化等方法对近 20 年生物样本库 SCIE 文献进行分析,探究其发展历程、国家与机构合作网络和研究热点演变情况,以期促进该领域研究持续推进。

2 资料与方法

2.1 资料来源

以 Web of Science 核心集合为数据来源,检索时间为 2020 年 7 月 26 日。以"biobank"为主题词

〔修回日期〕 2021-05-25

〔作者简介〕 高雅洁,硕士,助理馆员,发表论文 1 篇;通讯作者:蔡国龙,教授,发表论文 50 余篇,参编著作 5 部。

〔基金项目〕 浙江省医药卫生科技计划省部共建重大项目“组织灌注导向的重症患者预警与治疗技术研发及应用”(项目编号:WKJ-ZJ-2001);浙江省自然科学基金“IL-33 调控脓毒症巨噬细胞焦亡的机制研究”(项目编号:LY20H150007);浙江省科技厅公益技术应用研究计划“基于分级诊疗的 18 岁以上人群血压专业管理技术评价与应用研究”(项目编号:2016C33123)。

进行检索,限定 2000 – 2019 年关于生物样本库 SCIE 文献,得到 4 387 篇文献,排除会议、综述、报道及重复文献,最终有效文献 3 011 篇。

2.2 研究方法

在文献计量基础上绘制生物样本库知识图谱,通过 Excel 统计研究历程、国家和机构发文情况以及期刊载文量。运用 CiteSpace 对合作网络、文献共被引、关键词演化路径进行共现分析。利用 VOS Viewer 绘制主题聚类 and 密度知识图谱,以可视化方式揭示生物样本库研究热点。

3 研究结果

3.1 生物样本库研究阶段

从文献数量角度看,生物样本库研究可划分为 3 个阶段。第 1 阶段:2010 年以前为萌芽探索期。这一阶段对生物样本库领域研究较少,年载文量不足百篇,处于小范围理论探索阶段。第 2 阶段:2011 – 2015 年为平稳起步期。这一阶段开始关注生物样本库建设和发展,年载文量由 121 篇增长至 249 篇,呈缓慢增长趋势,总体发展平稳。第 3 阶段:2016 年至今为快速发展期。这一阶段生物样本库已成为学术界关注的热点话题,年载文量激增,生物样本库研究迎来快速发展。

3.2 合作关系网络

合作知识图谱展示合作密切程度和知识流动情况,不同国家/地区和机构间合作有利于知识共享和成果推进。欧洲各国地理位置近、欧盟紧密联系、交流合作密切,欧美国家对生物样本库关注程度较高,研究成果较多;亚洲载文量超过百篇的国家/地区仅有中国,发展相对缓慢,与其他国家合作较少。各机构合作网络体现较强的地域接近性,机构合作较多发生在同洲际之间。较欧美机构之间合作而言,跨洲际合作还亟待加强。

3.3 期刊分布与高被引文献分析

根据布拉德福定律,对主题相关文献的期刊按照数量递减排序划分出核心区、相关区以及边沿区^[4],核心区期刊具有较高发文量和发行量^[5]。载文量前 20 位的期刊平均影响因子 (IF) 为 5.784,生物样本库文献主要发表在遗传学、伦理学、生物分子学等学科期刊上。排名前 10 位的高被引文献中,发表在影响因子 (IF) >30 的高影响力期刊的占 40%。高被引文献代表在生物样本库研究领域国际期刊中被引用最多的文献,成为生物样本库研究奠基性文献。

3.4 聚类分析及热点演化

3.4.1 关键词聚类分析 关键词代表研究主题,对其聚类分析可知该领域研究热点。生物样本库相关研究关键词共现可视化图谱主要展示 4 个方向聚类:一是生物样本库,包括生物样本库自身建设、基因研究和生命伦理学;二是风险,通过研究疾病风险因素揭示生物样本库存在的意义;三是全基因组关联分析,通过对大规模群体全基因组分析揭示疾病发生、发展与治疗相关的遗传力;四是基因表达,主要研究生物样本库在肿瘤学中的应用价值,见图 1。生物样本库相关研究最热点之处是自身建设方面,涉及生命伦理、基因、遗传、疾病等。

3.4.2 研究热点演化路径 关键词时区图体现的是关键词首次出现的年份以及在时间轴上的变化,反映研究热点演进路径^[6],见图 2。2005 年前生物样本库发文量较少,主要研究热点是生物样本库自身建设与基础研究。2007 年后样本获得和存储的监管、伦理审查受到重视。2008 年生物样本库研究热度上升,基因表达和变异、疾病遗传风险因素等相关研究热度持续增高,队列研究成为热门研究方法。2010 年起研究热点更多关注疾病和质量监管。2014 年生物样本库数据共享首次被关注。2018 年后基因遗传风险评估作为疾病风险研究分支快速发展,精神类疾病研究

新热点。

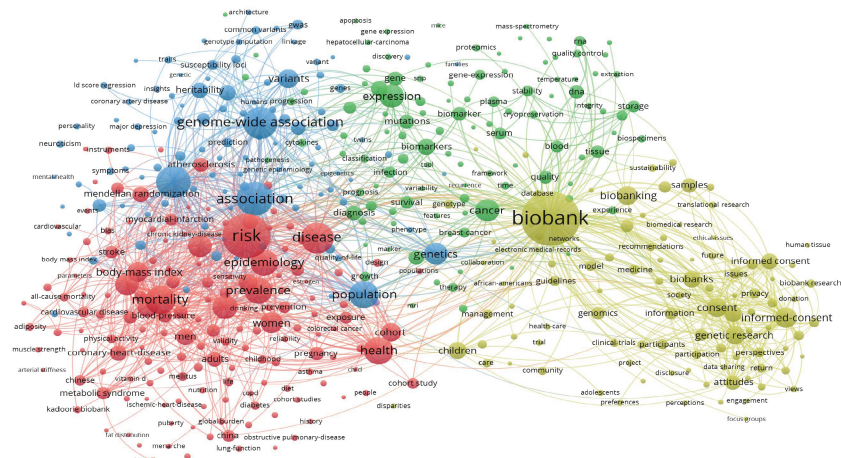


图 1 关键词共现聚类图谱

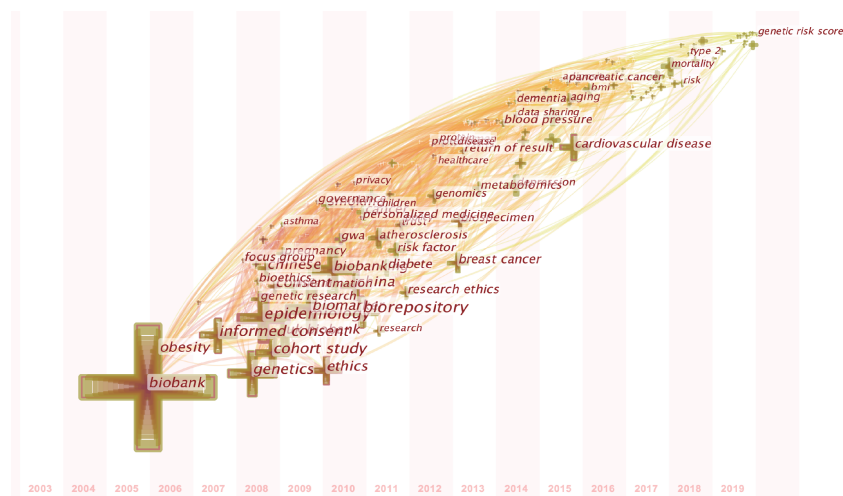


图2 关键词共现时区

4 讨论

4.1 完善生物样本库应用机制

从理论层面来看,生物样本库研究已经从数据获取和存储的基本建设方面转变为基因组关联分析研究上;从监管层面来看,对生物样本库标准化流程管理和隐私保护伦理审查日益被重视;从应用层面来看,生物样本库将会重点关注基因组学、代谢组学、生物标志物筛选研究,癌症、心血管病、糖尿病、抑郁症、阿尔兹海默病等遗传性疾病和精神类疾病是未来研究方向。生物样

本库的发展要充分考虑应用层面,建立符合自身特点的发展机制^[7],应通过激励政策建立和实施生物样本库发展应用机制,政府和行业协会等权威机构应在法律法规和指南制定方面发挥指导作用,明确样本应用规范。精准医学开启了利用大数据和信息化技术从基础研究向临床研究转化的时代^[8],将生物样本产生的生理生化信息和遗传组学信息与大数据处理技术有机结合,充分发挥生物样本库应用价值。

4.2 建立合作共赢的关系网

共享共赢才能建设可持续发展、研究价值最大

化的生物样本库^[9]。研究发现目前开展生物样本库研究的主要机构是高校、医院和科研院所,各机构合作网络显示出地域接近性,机构之间的合作多在同国和同洲际之间,亚洲国家和机构之间的合作较欧美国家而言亟待加强。生物样本库研究日益丰富,内容涉及遗传学、伦理学、生物分子学、公共卫生学等学科,其建设与发展涉及跨学科、跨专业、跨机构的协同合作。多学科交叉融合不仅需要扎实的专业知识,还应具备开展生物数据分析、多组学分析的能力。只有充分整合资源和共享数据才能更好地实现成果转化,建设规模化、规范化的生物样本库。

4.3 加强生物样本库标准化建设

随着生物样本库研究领域不断发展、文献量不断增加、学科不断扩大,对生物样本库相关政策法规、伦理审查和标准化管理提出更高要求。无论是生物样本采集与存储,或是对生物样本库进行二次使用的研究,都需要进行严格监管,符合伦理学“不伤害、公平和尊重”基本原则^[10]。由于缺乏统一标准,生物样本库建设和管理水平良莠不齐,异质性高的多源生物样本资源难以整合,阻碍资源共享应用^[2]。加强生物样本库标准化建设,应制定符合国情的质量管理体系和相关管理机制,完善顶层设计,发布配套标准,有法可依、有章可循,实现人员、样本、信息、机制闭环管理。只有在标准化流程监管下才会产生高质量的转化医学成果,实现精准医学模式实践。

4.4 推动生物样本库可持续性发展

生物样本库研究未进入可持续发展阶段^[11]。2008 年“风险”作为高频关键词出现在研究热点中,2018 年“基因遗传风险评分”作为疾病风险管控重要手段再次成为新的研究热点。可见重视生物样本规范化管理并加强科技创新对风险的管控,才能使生物样本库建设进入良性循环的发展状态,从而推动生物样本库可持续性发展。可从投入和产出两方面考虑,建立和完善良好的自身

造血机制,投入与产出缺一不可^[12]。投入一方面是从财政和政策制定角度加强支持与监管,另一方面是培养科技人才、提高公众认知水平;产出主要体现在临床科研向精准医学迈进。只有建立良好的投入、产出循环体系,才能从根本上实现生物样本库可持续性发展。

4.5 对中国生物样本库建设的启示

中国关于生物样本库研究的 SCIE 发文量居全球第 3 位,而排名前 10 位的高被引文献中无一来自中国。中国生物样本库建设起步时间晚,但发展速度快,国内生物样本库建设受到的关注度逐年提升,不同类型的样本库也应运而生。样本资源丰富,但实际利用率很低。因此以模式创新带动科技创新,以科技创新拉动服务创新,促进生物样本库应用转化显得尤为重要^[13]。中国要在生物样本库领域具有国际话语权,首先要建立长效发展机制,完善标准化管理制度,其次要抓住研究热点,提高科研质量,鼓励将研究成果发表在具有国际学术影响力的高质量期刊上,另外要重视科研成果转化与应用,优化对生物样本库资源利用,创新应用导向。

5 结语

本文通过对 Web of Science 核心集合中近 20 年关于生物样本库的 SCIE 文献进行知识图谱可视化分析,总结出生物样本库萌芽探索期、平稳起步期和快速发展期 3 个阶段;绘制合作网络知识图谱,欧美国家和机构对生物样本库关注程度更高、研究成果更多;高被引文献具有奠基性作用,对推进生物样本库研究可持续发展具有重要意义;通过关键词聚类共现图谱分析研究热点及演化路径。生物样本库发展至今,机遇与挑战并存,目前仍存在投入大、产出少,低层次重复建设、缺乏自身造血机制,空间、设备、人员和资金浪费的问题^[14]。未来生物样本库发展需要更加规范、更多产出、更优配置。

参考文献

- 董尔丹, 胡海, 俞文华. 生物样本库是生物医学研究的重要基础 [J]. 中国科学: 生命科学, 2015, 45 (4): 359–370.
- 满秋红, 王伟业. 重构生物样本库建设理念, 促进临床研究资源协同应用与发展 [J]. 协和医学杂志, 2019, 10 (4): 409–413.
- 崔晓英, 宋双, 翟倩雯, 等. 精准医学时代下中医特色生物资源样本库的建设探析 [J]. 江苏科技信息, 2020, 37 (9): 25–27, 40.
- 陈万思, 白雨菲. 新中国成立 70 年来劳动关系研究知识图谱与演进——基于 1949–2018 年间 5820 篇文献的计量分析 [J]. 中国劳动关系学院学报, 2020, 34 (3): 51–62, 84.
- 石晶晶, 石树青, 胡元会, 等. 心血管疾病能量代谢研究的可视化分析 [J]. 中国循证心血管医学杂志, 2020, 12 (4): 428–432, 442.
- 钱文静, 张有奎. 人类命运共同体研究的可视化知识图谱分析 [J]. 西南民族大学学报 (人文社科版), 2020, 41 (7): 222–230.
- 李海燕, 倪明宇, 林金嬉, 等. 临床生物样本库应用机制探索 [J]. 中华医学科研管理杂志, 2019, 32 (5): 397–400.
- 韩江, 周为民, 陈涵, 等. 如何发挥生物样本库在临床研究中的作用 [J]. 转化医学杂志, 2018, 7 (6): 346–349.
- 陈思静, 吴茂锋, 李佩娟. 生物样本库的建设与发展 [J]. 生物化工, 2019, 5 (4): 164–166.
- 赵励彦, 范贞, 刘瑞爽, 等. 生物样本库的伦理审查 [J]. 中国医学伦理学, 2020, 33 (3): 345–348.
- 雷瑞鹏, 冯君妍. 生物样本数据库的泡沫已经破裂了吗?——数字化时代对生物样本数据库可持续发展的挑战 [J]. 医学与哲学, 2018, 39 (9): 90–95.
- 满秋红, 于农, 闫飞, 等. 我国生物样本库建设现状与未来发展的思考 [J]. 中国医药生物技术, 2018, 13 (4): 289–293.
- 刘峙雅. 我国生物样本库的研究进展 [J]. 现代医药卫生, 2021, 37 (5): 759–763.
- 郜恒骏, 杜莉利, 张小燕, 等. 生物样本库发展的现状、机遇与挑战 [J]. 协和医学杂志, 2018, 9 (2): 172–176.
- Zhou X, Wu Z, Yin A, et al. Ontology Development for Unified Traditional Chinese Medical Language System [J]. Artificial Intelligence in Medicine, 2004, 32 (1): 15–27.
- 陈欢欢. 图书情报学领域本体的构建研究 [J]. 图书馆学研究, 2011 (11): 11–16.
- 郝伟学. 中医健康知识图谱的构建研究 [D]. 北京: 北京交通大学, 2017.
- 贾李蓉, 刘静, 于彤, 等. 中医药知识图谱构建 [J]. 医学信息学杂志, 2015, 36 (8): 51–53.
- 刘丽红, 刘静, 李海燕, 等. 《中国中医药学主题词表》“中药功能分类”主题词选词研究 [J]. 中国中医药信息杂志, 2015, 22 (10): 38–40.
- 肖洁, 冒宇清, 李新霞. 中医病证本体库的构建研究 [J]. 无线互联科技, 2018, 15 (13): 115–116, 133.
- 李明, 朱邦贤, 周强. 基于 protégé 的中医证候本体构建方法研究 [J]. 数理医药学杂志, 2015 (6): 807–809.
- 侯玉, 张昌林, 周强, 等. Protégé 在构建中医治则治法本体中的运用 [J]. 医学信息 (上旬刊), 2007, 20 (11): 1866–1871.
- 倪萍, 陆宇宏. 中医方剂领域本体建模和语义推理实现 [J]. 现代情报, 2012 (6): 133–140.
- Miao F, Liu H, Huang Y, et al. Construction of Semantic-based Traditional Chinese Medicine Prescription Knowledge Graph [C]. Chongqing: 2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), 2018.
- 刘丽红, 贾李蓉, 刘静, 等. 中医临床术语系统 v2.0 “中药饮片”类概念及关系设定研究 [J]. 中国中医药信息杂志, 2019, 26 (1): 95–98.
- 国家药典委员会. 中华人民共和国药典: 2020 版 [M]. 北京: 中国医药科技出版社, 2020.
- 朱彦, 贾李蓉, 高博, 等. 中医临床术语系统 v2.0 设计与构建 [J]. 中国中医药图书情报杂志, 2018, 42 (3): 10–15.

(上接第 42 页)