

生物信息学在我国医学研究生学位论文中的应用现状：基于VOSviewer的文献计量分析



刘均立, 姬子怡, 吕金字, 崔一鸣, 洪 健

暨南大学基础医学与公共卫生学院 (广州 510632)

【摘要】目的 分析中国医学研究生学位论文中生物信息学应用现状。**方法** 基于中国知网与万方数据库, 筛选 2003 年 1 月 1 日至 2023 年 12 月 31 日发表的生物信息学主题医学研究生学位论文。采用 VOSviewer 软件分析我国医学研究生的生物信息学研究热点, 并以肿瘤学领域为例, 分析生物信息学在临床实践中的应用现状。**结果** 共纳入学位论文 22 298 篇, 其中硕士学位论文 14 036 篇 (62.95%), 博士学位论文 8 262 篇 (37.05%)。20 年来, 我国医学研究生生物信息学主题学位论文数量呈持续增长态势, 研究热点主要集中在表达分析、机制研究、数据库与肿瘤学研究等方面。其中, 肿瘤相关研究占学位论文的 45.72%。以肺癌为例, 生物信息学为靶点探索、辅助诊断、药效监测等临床前研究提供了数据支持, 并已初步应用于临床诊断与辅助治疗决策。**结论** 生物信息学在我国医学研究生中的应用增长迅速, 尤其在肿瘤学领域, 生物信息学在其早期诊断、个体化治疗及预后评估中均发挥重要作用。

【关键词】 生物信息学; 肿瘤; 医学; 研究生; 文献计量学

【中图分类号】 G353.1; R73 **【文献标识码】** A

The application of bioinformatics on academic thesis of Chinese graduate student: a bibliometric analysis based on VOSviewer

LIU Junli, JI Ziyi, LYU Jinyu, CUI Yiming, HONG Jian

School of Medicine, Jinan University, Guangzhou 510632, China

Corresponding author: HONG Jian, Email: hongjian7@hotmail.com

【Abstract】Objective To analyze the application status of bioinformatics on Chinese medical graduate students. **Methods** Available academic thesis was filtered with a topic of bioinformatics medical research ranging from 2003 January 1st to 2023 December 31st in CNKI and WanFang Database. The VOSviewer software was used to analyze the research hotspots of the academic thesis of Chinese medical graduate students, taking oncology as an example, the application status of bioinformatics in clinical practice was investigated. **Results** A total of 22,298 theses were included, comprising 14,036 master's theses (62.95%) and 8,262 doctoral theses (37.05%). In the last two decades, the number of bioinformatics-related thesis for Chinese medical student has been increasing continuously. The research hotspots of academic thesis mainly focused on expression analysis, mechanistic study, database and oncology. Notably, oncology-related research accounted for 45.72% of the theses. For lung cancer, application of bioinformatics has provided data support for pharmacological target investigation, diagnosis assistance and therapeutic effect

DOI: 10.12173/j.issn.1004-5511.202408080

基金项目: 国家自然科学基金面上项目 (82473023)

通信作者: 洪健, 博士, 教授, 博士研究生导师, Email: hongjian7@hotmail.com

yxxz.whuzhmedj.com

monitoring in preclinical research. In addition, bioinformatics has been employed to aid clinical decision making, such as diagnosis and determination of therapeutic strategies. **Conclusion** The application of bioinformatics in the theses of medical postgraduates in China is increasing rapidly, particularly in the field of oncology. In clinical oncology practice, bioinformatics plays a pivotal role in early diagnosis, personalized therapy, and prognostic evaluation.

【Keywords】 Bioinformatics; Oncology; Medicine; Graduate student; Bibliometrics

生物信息学是随人类基因组计划启动而兴起的一门学科，其基于高通量分析手段可将大量生物数据转化为有效信息，基于大数据的生物信息学分析已在生命科学领域尤其是医学领域得到广泛应用。研究生是医学科研活动的重要力量，中国研究生培养单位中超过 60% 的基础研究由研究生参与完成^[1]；学位论文作为研究生科研成果的主要表现形式之一，能全面、系统地反映相应领域的研究现状^[2]。目前，医学研究生学位论文中的生物信息学应用现状较少报道，本文基于中国知网（CNKI）与万方数据库，对近二十年来（2003—2023 年）中国医学研究生的生物信息学主题学位论文进行梳理，分析生物信息学的研究热点，并以肿瘤学（肺癌）领域为例，探讨生物信息学在临床实践中的应用现状。

1 资料与方法

1.1 纳入与排除标准

纳入标准：医学领域生物信息学相关研究生学位论文。排除标准：①研究内容与主题不符；②信息或资料不全的论文；③不同数据库中重复纳入的论文。

1.2 文献检索策略

计算机检索 CNKI 与万方数据库收录的主题为生物信息学的研究生学位论文，检索时限为 2003 年 1 月 1 日至 2023 年 12 月 31 日。在检索结果中，以“医学”为学科进行筛选（万方数据库：选择“医药、卫生”类；中国知网：选择“肿瘤学、基础医学”等相关学科）。

1.3 文献筛选与资料提取

通过 EndNote X9 软件去除重复文献后，由 2 名研究者严格按照纳入与排除标准，逐一阅读论文的标题与摘要进行复筛，获得相关生物信息学主题学位论文。提取论文基本信息，包括文献标题、作者、作者学位类型、发表年份及研究对象等。

1.4 统计学分析

关键词是研究主题的高度概括，其出现频次代表了相关研究成果的数量及研究内容的集中程度，可反映论文的研究领域和研究方向^[3]。使用 VOSviewer 1.6.19 软件对关键词进行归类与可视化，节点体积的大小代表其出现频次，体积越大提示该关键词在领域内受关注度越高；不同节点之间的连线反映其关联性，连线粗细代表共现频次；相同颜色的关键词属于同一聚类，每个色系代表一个独立的研究主题集群。分析类型设置为关键词/共现，设定阈值为 5（过滤去除出现频次 ≤ 5 的关键词），选择出现频次前 1 000 的关键词构建共现网络图谱。

2 结果

2.1 年发文量情况

共纳入学位论文 22 298 篇，其中硕士学位论文 14 036 篇（62.95%），博士学位论文 8 262 篇（37.05%），学位论文数量总体上呈持续增长态势，年发文量从 2003 年的 123 篇逐渐增长至 2022 年的 2 669 篇。其中，硕士论文数量自 2004 年超过博士论文后持续维持较高增速（年均增速 25.06%）；而博士论文增长率波动较大，分别于 2008、2011、2014 与 2022 年出现负增长（年均增速 14.88%），论文占比由 2003 年的 52.03% 降至 2022 年的 30.12%，详见图 1。鉴于数据库收录论文存在延迟，本研究未对 2023 年发文量作进一步分析。

2.2 机构分布情况

纳入论文共涉及 112 所高校与科研机构。发文量排名前 10 的机构中 4 所为医科大学或独立医学院校，累计发文 2 598 篇（占总发文量的 11.65%）；余下 6 所为综合性院校（发文主要来自其附属医学院/医学部）。排名前 10 的机构累计发文 6 139 篇，即 8.93% 的机构发表了 27.53% 的学位论文，而 72.47% 的论文则由其余 102 个

机构发表, 详见表 1。

2.3 关键词共现情况

纳入论文共涉及 2 603 个关键词, 依据聚类结果及中心度情况, 如图 2 所示可归为 11 个聚类, 其中聚类 #1 (红色) 主要与表达相关, 多涉及基因表达及其功能与定位关系的研究、基因克隆与表达载体的构建; 聚类 #2 (绿色) 主要针对不同临床疾病机制的探索; 聚类 #3 (蓝色) 关注不同基因在肿瘤发生发展中的作用与调控机制, 以及靶向药物的筛选与疗效探索; 聚类 #4 (黄色) 主要关注遗传突变基因的筛选与致病性分析, 以及相关疾病的发病风险预测与基因治疗等; 聚类 #5 (紫色) 主要包括与恶性肿瘤预后及耐药相关的研究; 聚类 #6 (青色) 主要为肿瘤免疫相关研究; 聚类 #7 (橙色) 主要关注肿瘤生物标志物的筛选与功能分析。

依据所含关键词, 可将上述聚类的研究热点归纳为三个主题: ①表达分析: 聚类 #1, 研

究方向包括基因表达、基因敲除、蛋白质结构与功能预测等; ②致病机制研究: 聚类 #2、聚类 #4, 发掘潜在标志物, 探索其功能及作用机制, 并最终应用于疾病的临床诊疗; ③肿瘤学研究: 聚类 #3、聚类 #5、聚类 #6、聚类 #7, 以恶性肿瘤为主要研究对象, 着重探索恶性肿瘤的治疗、耐药、预后等方面。余下聚类包含关键词数量较少、聚类不明显 (分别为生物医学、标志物、网络分析与数学等相关研究), 故未对其作详细分析。

2.4 肿瘤领域生物信息学应用情况

选取肿瘤学领域, 分析生物信息学在临床及临床前研究的应用现状, 结果显示肿瘤学研究相关论文 10 194 篇, 占比 45.72%, 发文量排名前 10 的癌种如表 2 所示。

以肺癌尤其是晚期非小细胞肺癌为例, 生物信息学在其临床前研究与精准诊疗方面发挥关键作用, 如表 3 所示。

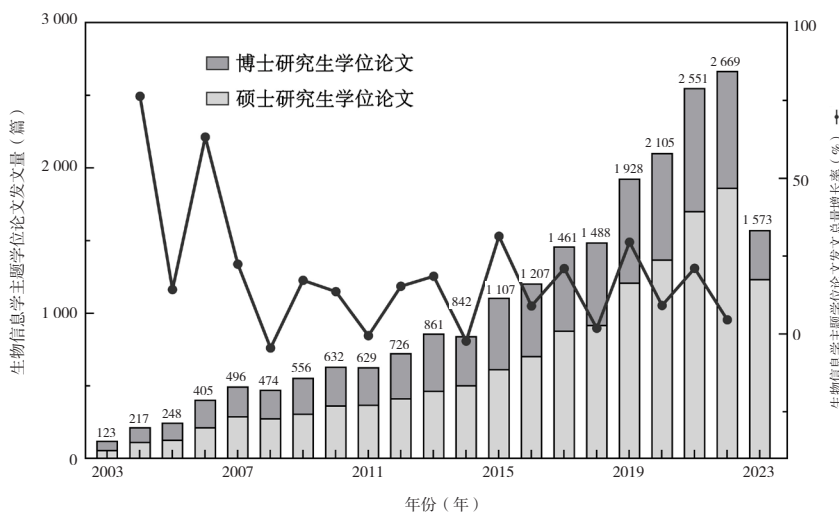


图1 2003—2023年生物信息学主题学位论文发文量变化趋势

Figure 1. Trends in the publication volume of bioinformatics-themed dissertations from 2003 to 2023

表1 生物信息学主题学位论文发文量前10机构

Table 1. The top 10 institutions publishing bioinformatics-related dissertations

| 排名 | 机构名称 | 硕士论文 | 博士论文 | 发文合计 | 百分比 (%) |
|----|--------|------|------|------|---------|
| 1 | 南方医科大学 | 451 | 395 | 846 | 3.79 |
| 2 | 郑州大学 | 473 | 302 | 775 | 3.48 |
| 3 | 山东大学 | 307 | 465 | 772 | 3.46 |
| 4 | 重庆医科大学 | 434 | 298 | 732 | 3.28 |
| 5 | 苏州大学 | 362 | 291 | 653 | 2.93 |
| 6 | 广西医科大学 | 413 | 170 | 583 | 2.61 |
| 7 | 浙江大学 | 174 | 298 | 472 | 2.12 |
| 8 | 天津医科大学 | 255 | 182 | 437 | 1.96 |
| 9 | 中南大学 | 192 | 243 | 435 | 1.95 |
| 10 | 华中科技大学 | 171 | 263 | 434 | 1.95 |

3 讨论

本研究对 CNKI、万方数据库 2003—2023 年的医学研究生学位论文进行了分析,结果显示,生物信息学相关论文发文量持续增长,侧面反映测序技术快速发展产生的大量数据亟需高效处理方式,生物信息学的强大数据处理能力成为其被广泛应用的基础。从文献发表单位可见,开展生物信息学研究的机构发文量分布较不均匀。以南方医科大学和郑州大学为代表的高校是生物信息学相关研究生的主要培养单位,在生物信息学研究方面较为活跃。但由于学位论文发文量还受培养单位招生人数等方面影响,若需准确探讨机构的科研实力和学术方向还需考虑更多因素。

关键词共现与聚类分析结果表明,基因表达分析、疾病机制研究以及恶性肿瘤的治疗、耐药、预后等为我国医学研究生的主要关注热点。目前多数医学研究集中于对某一致病基因、标志物蛋白的探索,通过基因敲除、基因克隆等表达调控技术结合生物信息学手段,在基因组、转录组或蛋白组层面进行表达分析,寻找差异表达的潜在靶点;进行功能注释和富集分析,分析其生物学过程并揭示分子机制,如探索细胞凋亡、氧化应激、炎症反应等相关通路在疾病中的表达情况^[12],以达到明确其功能、作用机制及临床价值的目的,是国内医学领域高效研究的模式之一。此外,聚类中基础实验相关关键词(“动物模型”“免疫组化”等)的相继出现提示了生物信息学分析与基础实验结合正逐渐成为我国医学研究生的普遍研究模式。其中,恶性肿瘤是全球范围内威胁人类健康的主要疾病之一,其本质上是由基因突变或表达异常,导致细胞增殖失控并最终形成恶性表型^[13],具有治疗需求迫切、研究基数庞大、复杂性(高度异质性、复杂分子机制和动态演化特性等)远超其他疾病等特点,而生物信息学凭借基因组分析、多组学整合和高通量数据分析能力,成为了恶性肿瘤科研与临床的核心助力平台,并促使肿瘤学成为生物信息学最热门的应用领域之一。

在恶性肿瘤的早期发现、精准诊断与个性化临床治疗中,生物信息学发挥着重要作用^[14]。如通过高通量测序技术获取患者样本和正常组织的基因组数据,识别具有临床意义的分子靶标,为

诊断标志物和治疗靶点的发现提供线索;使用计算机辅助药物设计技术,预测药物与靶点的相互作用,筛选出有效候选药物或优化已有药物方案;基于不同的诊断标志物(如 miRNA、ctDNA、外泌体等),实现恶性肿瘤早期筛查;或提取患者的基因表达谱,为其制定个体化的诊疗方案^[15]。随着人工智能和机器学习在生物信息学中的应用,未来临床生物信息学有望帮助医生更加精准地诊断疾病、预测预后,并为患者量身定制个性化治疗方案。

综上所述,生物信息学在医学研究生科研中的应用不断增长,表达分析与机制研究作为连接基础医学和临床的桥梁,推动着医学研究向临床应用的转化,尤其是在肿瘤学研究领域,而生物信息学亦是实现肿瘤早期诊断、治疗及疗效预测的有效手段。

伦理声明: 不适用

作者贡献: 研究设计:洪健、刘均立;数据收集与分析:刘均立、姬子怡、吕金宇、崔一鸣;论文撰写:刘均立、姬子怡;论文审阅:洪健

数据获取: 本研究中使用和(或)分析的所有数据均包含在本文中

利益冲突声明: 无

致谢: 不适用

参考文献

- 1 王战军,张微.70年探索奋斗:中国研究生教育发展规律与启示[J].学位与研究生教育,2019,(9):43-48.[Wang ZJ, Zhang W. 70 years exploration of the norm of graduate education in China[J]. Academic Degrees & Graduate Education, 2019, (9): 43-48.] DOI: [10.16750/j.adge.2019.09.008](https://doi.org/10.16750/j.adge.2019.09.008).
- 2 曲继鹏.基于CNKI数据库的毕业设计教学文献分析[J].西昌学院学报(自然科学版),2019,33(3):124-128.[Qu JP. Analysis of graduation design teaching documents based on CNKI database[J]. Journal of Xichang University (Natural Science Edition), 2019, 33(3): 124-128.] DOI: [10.16104/j.issn.1673-1891.2019.03.029](https://doi.org/10.16104/j.issn.1673-1891.2019.03.029).
- 3 谢寒杰,晋治波,盖京苹.我国微生物学相关方向研究生论文选题变化分析(1999—2018年)[J].微生物学通报,2022,49(8):3539-3549.[Xie HJ, Jin ZB, Gai JP. Research topics in microbiology for graduate students in China (1999—2018)[J]. Microbiology China, 2022, 49(8): 3539-3549.] DOI: [10.13344/j.microbiol.china.211133](https://doi.org/10.13344/j.microbiol.china.211133).
- 4 Wang Z, Wu L, Li B, et al. Toripalimab plus chemotherapy for patients with treatment-naive advanced non-small-cell lung

- cancer: a multicenter randomized phase III trial (CHOICE-01)[J]. *J Clin Oncol*, 2023, 41(3): 651–663. DOI: [10.1200/jco.22.00727](https://doi.org/10.1200/jco.22.00727).
- 5 Chiappori AA, Creelan B, Tanvetyanon T, et al. Phase I study of taminadenant (PBF509/NIR178), an adenosine 2A receptor antagonist, with or without spartalizumab (PDR001), in patients with advanced non-small cell lung cancer[J]. *Clin Cancer Res*, 2022, 28(11): 2313–2320. DOI: [10.1158/1078-0432.Ccr-21-2742](https://doi.org/10.1158/1078-0432.Ccr-21-2742).
- 6 Cooper AJ, Kobayashi Y, Kim D, et al. Identification of a ras-activating TMEM87A-RASGRF1 fusion in an exceptional responder to sunitinib with non-small cell lung cancer[J]. *Clin Cancer Res*, 2020, 26(15): 4072–4079. DOI: [10.1158/1078-0432.Ccr-20-0397](https://doi.org/10.1158/1078-0432.Ccr-20-0397).
- 7 Chaft JE, Oezkan F, Kris MG, et al. Neoadjuvant atezolizumab for resectable non-small cell lung cancer: an open-label, single-arm phase II trial[J]. *Nat Med*, 2022, 28(10): 2155–2161. DOI: [10.1038/s41591-022-01962-5](https://doi.org/10.1038/s41591-022-01962-5).
- 8 Owonikoko TK, Park K, Govindan R, et al. Nivolumab and ipilimumab as maintenance therapy in extensive-disease small-cell lung cancer: CheckMate 451[J]. *J Clin Oncol*, 2021, 39(12): 1349–1359. DOI: [10.1200/jco.20.02212](https://doi.org/10.1200/jco.20.02212).
- 9 Nakagawa K, Garon EB, Seto T, et al. Ramucirumab plus erlotinib in patients with untreated, EGFR-mutated, advanced non-small-cell lung cancer (RELAY): a randomised, double-blind, placebo-controlled, phase 3 trial[J]. *Lancet Oncol*, 2019, 20(12): 1655–1669. DOI: [10.1016/s1470-2045\(19\)30634-5](https://doi.org/10.1016/s1470-2045(19)30634-5).
- 10 Sullivan FM, Mair FS, Anderson W, et al. Earlier diagnosis of lung cancer in a randomised trial of an autoantibody blood test followed by imaging[J]. *Eur Respir J*, 2021, 57(1): 2000670. DOI: [10.1183/13993003.00670-2020](https://doi.org/10.1183/13993003.00670-2020).
- 11 Abbosh C, Frankell AM, Harrison T, et al. Tracking early lung cancer metastatic dissemination in TRACERx using ctDNA[J]. *Nature*, 2023, 616(7957): 553–562. DOI: [10.1038/s41586-023-05776-4](https://doi.org/10.1038/s41586-023-05776-4).
- 12 周理韡, 郑书栋, 黄硕, 等. 生物信息技术在生殖领域中的研究进展[J]. *国际生殖健康 / 计划生育杂志*, 2022, 41(6): 487–493. [Zhou LW, Zheng SD, Huang S, et al Research progress of bioinformatics technology in reproductive system[J]. *Journal of International Reproductive Health/Family Planning*, 2022, 41(6): 487–493.] DOI: [10.12280/gjszjk.20220361](https://doi.org/10.12280/gjszjk.20220361).
- 13 李伟. 生物信息学分析对于肿瘤学硕士研究生培养的意义[J]. *中国继续医学教育*, 2023, 15(10): 170–173. [Li W. Significance of bioinformatics analysis for the training of postgraduates in oncology[J]. *China Continuing Medical Education*, 2023, 15(10): 170–173.] DOI: [10.3969/j.issn.1674-9308.2023.10.038](https://doi.org/10.3969/j.issn.1674-9308.2023.10.038).
- 14 中国抗癌协会肿瘤基因诊断专业委员会, 中国抗癌协会整合肿瘤学分会. 肿瘤基因诊断二代测序生物信息学分析平台本地化建设中国专家共识(2024年版)[J]. *中国癌症防治杂志*, 2024, 16(6): 638–649. [Tumor Gene Diagnosis Committee of the China Anti-Cancer Association, Integrative Oncology Branch of the China Anti-Cancer Association. Chinese expert consensus on the localization of bioinformatics analysis for tumor genetic diagnosis using next-generation sequencing (2024 edition)[J]. *Chinese Journal of Oncology Prevention and Treatment*, 2024, 16(6): 638–649.] DOI: [10.3969/j.issn.1674-5671.2024.06.02](https://doi.org/10.3969/j.issn.1674-5671.2024.06.02).
- 15 武文葵. 生物信息学在生物医学领域的应用[J]. *学园*, 2024, 17(23): 95–97. [Wu WY Applications of bioinformatics in the biomedical field [J]. *Academy*, 2024, 17(23): 95–97.] <https://www.cnki.com.cn/Article/CJFDTotol-XYJK202423031.htm>

收稿日期: 2024 年 08 月 28 日 修回日期: 2025 年 05 月 11 日

本文编辑: 桂裕亮 曹越

引用本文: 刘均立, 姬子怡, 吕金宇, 等. 生物信息学在我国医学研究生学位论文中的应用现状: 基于VOSviewer的文献计量分析[J]. *医学新知*, 2025, 35(12): 1501–1506. DOI: [10.12173/j.issn.1004-5511.202408080](https://doi.org/10.12173/j.issn.1004-5511.202408080).

Liu JL, Ji ZY, Lyu JY, et al. The application of bioinformatics on academic thesis of Chinese graduate student: a bibliometric analysis based on VOSviewer[J]. *Yixue Xinzhi Zazhi*, 2025, 35(12): 1501–1506. DOI: [10.12173/j.issn.1004-5511.202408080](https://doi.org/10.12173/j.issn.1004-5511.202408080).