

# 肿瘤内科学专业学位论文研究生医学公共大数据运用能力培养

贾馨竹<sup>1,2</sup> 高伊星<sup>1,2</sup> 张岸梅<sup>1,2</sup>

<sup>1</sup>陆军军医大学第二附属医院肿瘤科,重庆 400037; <sup>2</sup>重庆市免疫治疗重点实验室 400037

通信作者:张岸梅,Email:8015003@qq.com

**【摘要】** 目前,探索性思维和利用医学公共大数据能力的培养易被忽视,导致肿瘤内科学专业学位论文研究生在科研实践中利用公共大数据发掘、证实临床新现象、新规律的能力明显不足。本研究通过介绍 TCGA 等常用医学公共大数据库的访问与下载工具,引入重复经典分析案例的学习方法,提出利用公共大数据验证临床猜想的研究方法。期望为提高肿瘤内科学专业学位论文研究生发现临床问题、总结科学规律的研究能力提供有益启示和参考。

**【关键词】** “双轨合一”培养模式; 肿瘤学专业学位论文研究生; 医学公共大数据; 生物信息学

**【中图分类号】** R73-31

DOI:10.3760/cma.j.cn116021-20191204-00125

## Practice of cultivating the ability of big data mining in graduates working for professional degree in medical oncology

Jia Qingzhu<sup>1,2</sup>, Gao Yixing<sup>1,2</sup>, Zhang Anmei<sup>1,2</sup>

<sup>1</sup>Institute of Cancer, Xinqiao Hospital, The Army Medical University, Chongqing 400037, China; <sup>2</sup>Chongqing Key Laboratory of Immunotherapy, Chongqing 400037, China

Corresponding author: Zhang Anmei, Email: 8015003@qq.com

**【Abstract】** At present, the cultivation of exploratory thinking and the ability to use medical public big data is easy to be ignored, which leads to the obvious lack of the ability of graduate students in oncology to use public big data to discover and confirm new clinical phenomena and new laws in scientific research. This study introduces the access and download tools of common medical big databases such as TCGA, introduces the learning method of repeated classic analysis cases, and proposes research methods to verify clinical hypothesis using public big data. It is expected to provide useful enlightenment and reference for improving the research ability of graduate students in oncology specialty to find clinical problems and summarize scientific laws.

**【Key words】** Double-track integrated training mode; Graduate working for professional degree in medical oncology; Publicly accessible medical big data; Bioinformatics

DOI:10.3760/cma.j.cn116021-20191204-00125

伴随着我国临床医学研究生培养目标由学术能力为牵引向以综合能力为牵引的战略性转型,掀起了专业学位研究生培养与住院医师规范化培训“双轨合一”模式的探索性改革<sup>[1]</sup>。“双轨合一”这一

创新培养模式解决了课堂教育、实验室训练与临床从业能力培训脱节的问题,逐步形成了一系列具有创新性、实践性和示范性的宝贵经验<sup>[2]</sup>。但同时也发现,因培养时间紧凑、培养重点仍倾向于临床技能

的提高等原因,导致肿瘤内科学专业学位论文研究生开展临床研究的能力较弱,这也成为自身成长道路上难以逾越的障碍。

近年来,基于公共医学大数据的循证医学证据在总结疾病规律、制定诊疗规范中的作用逐渐凸显<sup>[3]</sup>。同时,以计算机技术为基础的生物信息学迅猛发展,为医学大数据的有效利用提供了保障<sup>[4]</sup>。在此背景下,本文基于本院“双规合一”培养模式大框架,分析了生物信息学在肿瘤内科学专业学位论文研究生培养中的重要性,就培养“公共大数据获取—生物信息学挖掘—临床新规律发现”能力的方法与步骤进行探讨。期望为当前“双轨合一”培养模式下进一步提高肿瘤内科学专业学位论文研究生发现临床问题、总结科学规律的研究能力提供有益的启示和参考。

## 1 加强肿瘤内科学专业学位论文研究生医学大数据利用能力的重要性

当前,肿瘤研究领域正经历着由经验医学向循证医学,再到精准医学的重大变革<sup>[5]</sup>。得益于信息化医疗手段的广泛应用,医疗实践中的人群预防、筛查诊断、干预治疗与预后随访等全要素过程无不伴随着数据的生成、采集与处理。此外,来自基础医学研究领域的基因组信息、转录组信息、生物标志物等同时积累了大量的数据,为转化医学的快速发展提供了可能<sup>[6]</sup>。而正是这些海量医学大数据的不断积累构成了推动精准医学发展的最基础要素。对医学大数据的合理使用,在海量的数据中发现潜在规律,验证并解决临床实践中观察到的问题,利用数据挖掘所得到的新规律、新知识指导临床诊疗手段的更新等问题,均需要多学科、多领域的共同参与来实现<sup>[7]</sup>。这也决定了具有临床医学、生物学、遗传学、统计学、计算机科学等相关知识背景的交叉型人才,未来必将在肿瘤精准治疗领域中发挥不可替代的作用。因此,接触医学大数据相关知识,掌握必要的方法与工具,实现对医学大数据的高效检索、整理、提取与挖掘,将是精准医学时代推动肿瘤疾病诊断与治疗水平发展的关键,也是肿瘤内科学专业学位论文研究生在科学研究的道路上实现自我提高的有力抓手。

区别于传统的基础和临床医学研究等统计学

分析任务,肿瘤医学大数据资源通常具有数据总量大、数据形式多样、整理规范性较低、单位数据中有效信息含量较少、生成更新快、数据质量参差不齐、冗余重复较普遍等几个显著特征。这就要求从事医学大数据分析的人员要掌握专业的分析方法与技能<sup>[8]</sup>。生物信息学是一门依托公共数据库,对生物医学信息进行获取、存储、检索和分析,在混杂信息中抽象和理解生物系统内在联系与规律的科学。具体来讲涉及数据的规范化存储与获取、序列比对、基因识别分析、分析遗传进化、序列重叠群装配、药物设计、生物系统等多个研究方向。目前,生物信息学已是我校临床医学专业学生的必修课。与其他课程相比,具有原理抽象、软件复杂、知识交叉面广等特点,被学生认为是较为困难的一门课程。但是,初步了解掌握一定的生物信息学知识对综合利用生物大数据,在大量而复杂的信息中揭示医学奥秘有极大的帮助,对提高人才科研素养和综合竞争力具有十分重要的作用。

近年来医学教育由传统的临床能力培养逐渐转变为临床与科研实力并重的新型教育模式。“公共大数据获取—生物信息学挖掘—临床新规律发现”能力的培养已经成为肿瘤内科学专业学位论文研究生未来发展不可或缺的一部分,有望成为我国当前“双轨合一”专业学位研究生培养模式的有益补充。培养适应我国人才成长需求的科学研究能力能够有效弥补我国肿瘤内科学专业学位论文研究生在当前培养模式所存在的不足,具有较强的紧迫性和现实意义。

## 2 “公共大数据获取—生物信息学挖掘—临床新规律发现”能力的培养

目前,研究生的临床工作安排通常较为紧凑,因而不适宜采取集中开设长学时课程的形式,而是建议在研究生培养全程中以已有的大数据挖掘得到循证医学证据为例,感受生物信息学方法的基本原理和应用方向,熟悉自身研究领域内公共大数据的获取渠道,并尝试重复现有的分析流程和结果。首先对使用生物信息学工具挖掘公共数据总结临床规律的全过程构建起直观的概念。进一步以此经验为基础,针对临床决策中遇到的困难,提炼出合理的科学问题与结论假设;并以问题为牵引寻找满

足要求的新的数据源与分析工具,在原有基础上不断扩展自身的能力范围,使肿瘤内科学专业学位论文在提升临床技能水平的同时具备进一步解决临床问题的能力。具体的培养过程应当包括如下三个阶段。

### 2.1 初步了解肿瘤学领域常用公共数据网站及分析工具

在常规肿瘤内科学教临床培训之外,应丰富以生物信息学基础知识为主的自学课程模块,内容包括回顾统计学基本知识,如连续变量检验分析、卡方检验、生存分析、Cox 风险分析等<sup>[9]</sup>;了解 The Cancer Genome Atlas 数据库、cBioPortal 数据库、SEER 临床信息数据库等的访问途径及数据类型;了解 TCGAblinks、TCGA Assembler、TCGA2stat、Firehose-FirebrowserR、RTCGA Toolbox、cBioPortal CGDS-R 等主流的第三方访问与下载工具<sup>[10]</sup>;通过完成工具软件内置的演示数据及分析流程来进行训练并记录成绩。

### 2.2 依托经典分析案例的模拟训练

包括既往经典的利用公共大数据进行回顾性分析得到的临床决策依据,以及最新发表的本领域的前沿案例,以 PBL 教学模式联合实际上机操作为核心,以经典分析案例提供的知识背景、数据来源及分析工具为线索,对其结论进行重复性、验证性分析。例如分析学习过程如下:既往研究发现,TTN 基因突变阳性可以作为晚期实体肿瘤患者接受免疫检查点抑制剂治疗中优势患者筛选的生物标志物,其机制是 TTN 阳性患者肿瘤突变负荷高。可通过 TCGA 数据库整理网站 cBioportal 中 27 种实体肿瘤、6 565 例患者的转录组(RNA-sequencing)数据、全外显子组数据(Wholeexomesequencing, WES)以及临床数据,得到 TTN 突变型患者的肿瘤突变负荷数据,即可验证 TTN 基因突变患者具有更高的肿瘤突变负荷;进而下载 Snyder 2014 黑色素瘤队列、vanAllen2015 黑色素瘤队列、Hugo2016 黑色素瘤队列、Riaz2017 黑色素瘤队列、Rizvi2015 非小细胞肺癌队列、Hellmann2018 非鳞非小细胞肺癌队列等临床治疗队列的 WES 及临床随访资料,验证 TTN 基因突变状态对患者接受免疫检查点抑制剂治疗后客观缓解率及生存时间的相关性,从而利用公共资源验证现有经典研究。通过对数个高水平经典案例

的学习研究及讨论,将前期接触的基础知识运用到解决具体问题的过程中来,逐步实现理论知识与实践操作的有机结合,激发学生的学习兴趣与主观能动性,持续提升利用生物信息学工具开展医学研究的能力。

### 2.3 临床观察的提炼与求证

在研究生主攻的课题方向上,根据自身的临床观察与体会,提出新的科学假设与推断,并整理得到清晰的实验设计、入组人群、干预条件、观察或评价指标、终点事件等,尝试通过检索医学公共大数据给出回顾性支持证据,为证实临床猜想提供依据。以抗原嵌合 T 细胞受体细胞(chimeric antigen receptor T-cell, CAR-T cells)过继回输治疗为例,临床实践中发现治疗前外周血中某细胞亚群与 B 细胞非霍奇金淋巴瘤患者治疗后无进展生存时间相关。为验证这一观察,可通过搜索 GEO 数据库中 CAR-T 细胞治疗公共队列,纳入同时记录有疾病诊断、基线人口学资料、生存时间随访及治疗前外周血组学数据的有效队列,比较该细胞群体与治疗有效性的关系,从而验证自身的临床观察是否具有普遍性,或对自身家属进行修正并进一步验证。在研究过程中,应为研究生提供必要的理论指导与个体化的帮助,其目的不在于掌握更多方法与技巧,而是让学生切实在实践的过程中灵活、准确选择正确的技术路线,最终解决复杂的科学问题。结合专业背景对统计分析结果进行解读,阐释其提示的医学规律,为进一步完善课题提供重要的数据支撑,让医学大数据与生物信息学工具真正成为肿瘤内科学专业学位论文科研道路上披荆斩棘的有力武器。

## 3 结语

“双轨合一”教学模式下专业学位研究生能力培养的核心是临床思维与实践技能的培养与提高。而对于肿瘤内科学专业学位论文研究生,不仅要强调临床水平的培养,同时应重视学生未来从事转化研究的能力的训练与潜力的提高。在生物信息学技术方法迅猛发展的今天,利用医学公共大数据的能力培养对于肿瘤专业研究生未来发科研道路发展起到举足轻重的作用。我国肿瘤内科学专业学位论文研究生医学教育需要顺应国际肿瘤学领域内最新的发展

趋势,在研究生“双轨合一”培养模式的基础上,着重培养具有“公共大数据获取—生物信息学挖掘—临床新规律发现”能力素养的肿瘤内科学专业执业医师;使其不仅具有高尚的医德医风、较高的临床技能水平与较丰富的初步经验积累,同时又具有最前沿、高效研究工具的使用经验;部分克服了临床研究过程中病历资料整理收集过程冗长所造成的困难,解决了当前科研能力培训中理论方法与现实需求连接不紧密、需求不清晰的问题,使肿瘤内科学专业学位研究生成为具有严密临床科研思维和较强科研发展前景、与国际医疗发展趋势接轨的复合型医学人才。

**利益冲突** 所有作者均声明不存在利益冲突

**作者贡献声明** 贾罄竹、高伊星,撰写论文;张岸梅,总体把关、审定论文

#### 参考文献

- [1] 韩莹. 临床医学专业学位研究生“双轨合一”培养模式的探讨[J]. 中国校外教育, 2016(9): 26-27. DOI: 10.3969/j.issn.1004-8502.2016.09.018.  
Han Y. Discussion on the "Dual-track Integration" training mode for clinical medical degree graduate students [J]. China After School Education, 2016(9): 26-27. DOI: 10.3969/j.issn.1004-8502.2016.09.018.
- [2] 郑璐,唐艺宸,谷慧英,等. 基于双轨合一模式的外科专业学位研究生“临床医学培训—医学统计学分析”一体化人才培养新模式[J]. 中国医学教育技术, 2018, 32(3): 249-251. DOI: 10.13566/j.cnki.cmet.cn61-1317/g4.201803004.  
Zheng L, Tang YC, Gu HY, et al. The new training mode of integrated clinical medical training and medical statistics analysis for clinical graduate education under the "Double-track Integration" training mode [J]. China Medical Education Technology, 2018, 32(3): 249-251. DOI: 10.13566/j.cnki.cmet.cn61-1317/g4.201803004.
- [3] 弓孟春,陆亮. 医学大数据研究进展及应用前景[J]. 医学信息学杂志, 2016, 37(2): 9-15. DOI: 10.3969/j.issn.1673-6036.2016.02.002.  
Gong MC, Lu L. Research progress and application prospect of medicalbig data [J]. Journal of Medical Intelligence, 2016, 37(2): 9-15. DOI: 10.3969/j.issn.1673-6036.2016.02.002.
- [4] 赵屹,谷瑞升,杜生明. 生物信息学研究现状及发展趋势[J]. 医学信息学杂志, 2012, 33(5): 2-6.  
Zhao Y, Gu RS, Du SM. The research status and development tendency of bioinformatics [J]. Journal of Medical Informatics, 2012, 33(5): 2-6.
- [5] 付文华,钱海利,詹启敏. 中国精准医学发展的需求和任务[J]. 中国生化药物杂志, 2016, 36(4): 1-4. DOI: 10.3969/j.issn.1005-1678.2016.04.01.  
Fu WH, Qian HL, Zhan QM. The needs and tasks of precision medicine development in China [J]. Chinese Journal of Biochemical Pharmaceutics, 2016, 36(4): 1-4. DOI: 10.3969/j.issn.1005-1678.2016.04.01.
- [6] 陈长仁,何发忠,周宏灏,等. 精准医学的基础研究与临床转化[J]. 中国药理学通报, 2015(12): 1629-1632. DOI:10.3969/j.issn.1001-1978.2015.12.001.  
Chen CR, He FZ, Zhou HH, et al. Basic research and clinical translation of precision medicine [J]. Chinese Pharmacological Bulletin, 2015(12): 1629-1632. DOI: 10.3969/j.issn.1001-1978.2015.12.001.
- [7] 王朝霞. 生物信息学:一门前沿交叉学科[J]. 安徽教育学院学报, 2002, 20(6): 56-57. DOI: 10.3969/j.issn.1674-2273.2002.06.020.  
Wang ZX. Bioinformatics: a cutting-edge interdisciplinary [J]. Journal of Anhui Institute of Education, 2002, 20(6): 56-57. DOI: 10.3969/j.issn.1674-2273.2002.06.020.
- [8] 姜鑫. 生物信息学数据库及其利用方法[J]. 现代情报, 2005, 25(6): 185-187. DOI: 10.3969/j.issn.1008-0821.2005.06.073.  
Jiang X. Bioinformatics database and its using method [J]. Modern Information, 2005, 25(6): 185-187. DOI: 10.3969/j.issn.1008-0821.2005.06.073.
- [9] 王倩. 加强统计学在临床科研中的应用[J]. 广东医学, 1999(3): 161-163.  
Wang Q. Strengthen the application of statistics in clinical research [J]. Guangdong Medical Journal, 1999(3): 161-163.
- [10] 谢龙祥,闫中义,党艺方. TCGA 数据库:海量癌症数据的源泉[J]. 河南大学学报(医学版), 2018, 37(3): 76-81.  
Xie LX, Yan ZY, Dang YF. TCGA database: the source of massive cancer data [J]. Journal of Henan University (Medical Edition), 2018, 37(3): 76-81.

(收稿日期:2019-12-04)

(本文编辑:唐宗顺)