

大数据时代下肿瘤病理学研究生科研能力培养模式探讨

罗韬, 姚小红, 时雨, 阎晓初, 卞修武

作者单位: 陆军军医大学(第三军医大学)第一附属医院病理学研究所, 重庆 400038

通信作者: 姚小红, 女, 副教授, 副主任医师, 硕士生导师, 研究方向为神经系统肿瘤病理诊断, E-mail: yxh15@163.com

基金项目: 重庆市研究生教育教学改革项目(YJG153072)

摘要: 大数据时代的到来, 对肿瘤病理学的研究生培养模式提出了新的挑战。一方面, 基因表达数据库, 文献共享, 微信等新一代互联网平台使研究生的科研工作可以快速入门; 另一方面, 也给研究生科研能力的培养带来了新的挑战。面对日新月异的大数据时代, 唯有努力提高理论和实践水平, 大力加强计算机应用能力, 始终强调逻辑思辨能力培养, 才能培养出适应新时代需要的肿瘤病理学人才。

关键词: 教育, 医学, 研究生; 病理学; 计算机通信网络; 信息系统; 大数据; 科研能力

Research on the scientific research ability training mode of tumor pathology graduate students in the era of big data

LUO Tao, YAO Xiaohong, SHI Yu, YAN Xiaochu, BIAN Xiuwu

Author Affiliation: Institute of Pathology, Southwest Hospital, the First Affiliated Hospital of Army Military Medical University (Third Military Medical University), Chongqing 400038, China

Abstract: The arrival of the big data era has proposed new challenges to the graduate students training model of tumor pathology. On the one hand, gene expression database, document sharing, WeChat and other new Internet platform enable graduate students to start their research quickly. On the other hand, it also brings new challenges to the cultivation of postgraduates' scientific research ability. In the ever-changing era of big data, we must constantly improve our level of theory and practice, focus on strengthening computer skills and always emphasize the cultivation of logical capability, to meet the requirement of new era.

Key words: Education, medical, graduate; Pathology; Computer communication networks; Information systems; Big data; Scientific research ability

病理学是基础医学与临床医学的桥梁学科, 研究疾病发生发展的过程和原理。肿瘤病理学, 主要研究肿瘤发生发展的过程, 吸引了包括病理学, 肿瘤学, 分子生物学, 遗传学, 信息学等多学科人才投入到这一领域, 为肿瘤病理学的发展带来了日新月异的变化^[1]。大数据(Big Data)又称巨量数据, 指的是所涉及的数据资料量规模巨大到无法通过人脑甚至主流软件工具进行处理以及分析。在生物医学领域, 主要包括存储在互联网数据库的包括DNA, RNA, 蛋白等海量数据, 也泛指相关的文献, 图书, 文献, 软件等资源^[2-3]。精准医学时代下的肿瘤学教育工作也越来越受到重视^[4]。如何分析、利用大数据时代下海量的生物医学数据, 培养肿瘤病理学研究生的科研能力, 是研究生导师, 带教教员面临的新的课题。

1 大数据时代信息传递/数据共享越加便利, 为研究生培养提供方便

随着互联网的发展, 越来越多以往无法获取的

资源变得随处可见, 这些资源的出现, 使研究生可以自行学习基本的实验技能^[5]。导师不需要直接带教, 而是把精力放在解决整理归纳研究生培养需要的资源以及解惑上, 这样可以有效提高研究生培养的效率和质量。

1.1 第三方数据平台助力研究课题 生物大数据时代的标志性成果就是第三方数据平台如雨后春笋般的出现。第三方数据平台主要指国内外学术/商业机构针对相应领域的研究成果, 为了方便科研工作者的需要, 在互联网搭建的各类生物学数据库。在肿瘤研究领域, 除PUBMED这类经典的数据库, 已涌现出一大批优质的数据库资源: 包括整合数据库(Genecard, Protein Atlas等), 泛肿瘤数据库(Oncomine^[6], TCGA^[7], Prognoscan等)以及特定肿瘤数据库(Gliovis, KM plotter等)等。下面分别介绍几个代表性的第三方数据库及其在肿瘤病理学研究生培养中的作用。

1.1.1 Genecards 这是一个内容丰富的基因综合数据库,不止提供包括基因,转录本,编码蛋白等基本信息,还包括指向GO与KEGG数据库的基因功能,人不同器官基因表达,蛋白质与蛋白质相互作用,抗体,重组蛋白和RNAi试剂等^[8]。对于研究生,可以通过此数据库“一站式”了解某分子的基本生物学功能,亚细胞定位,蛋白分子量,相互作用蛋白,适用抗体等重要信息。此外,该数据库直接链接了包括PUBMED, Swiss-Port, UCSC, dbSNP等一百余个公共数据库,可以直接链接到引用数据库查询详细信息。Genecards可以说是对基因的文献及数据库信息进行二次加工,极大提高科研效率。

1.1.2 Protein Atlas 肿瘤病理学的研究,除了新兴的基因突变指标,目前应用最为广泛的就是免疫组化。而Protein Atlas则是专门针对肿瘤分子蛋白染色的数据库。其包括约24 000种人类蛋白在48种人类正常组织,20余种肿瘤组织的免疫组化染色结果,并经专业人员阅读和标引,保证染色结果具有充分的代表性;并利用47个细胞系进行荧光染色,利用激光共聚焦显微镜显示该蛋白的亚细胞定位^[9]。2017年,该数据库与TCGA数据组合作,整合TCGA肿瘤/癌旁样本已有RNA-seq数据及预后等信息,并利用相应病人石蜡样本制作组织芯片,推出the Pathology Atlas计划,以蛋白组学的方式揭露肿瘤靶点^[10]。

1.1.3 Gliovis 除了各类泛肿瘤数据库外,还有不少以特定肿瘤为研究对象的数据库。其中最具代表性的是Gliovis。该数据库不仅整合了TCGA中胶质瘤的临床,预后,病理及分子分型等相关信息,还包括了其他胶质瘤研究数据,包括Rembrandt, Grevendeel等大型临床样本数据集,并根据文献研究进一步更新,为基础研究的第三方临床验证提供高质量的数据。此外,数据库还提供了多种统计学分析方法,如相关性分析,配对检验,整合分析等^[11]。利用该数据库进行多个样本集的综合分析,寻找在胶质瘤发病及恶性转化中的关键基因,培养研究生形成整合分析的思维方式。

1.2 文献下载更为便捷 虽然第三方数据库囊括了海量的基因及肿瘤信息,但依然存在两个重要问题:(1)数据库都需要人力进行整理,在即时性上会较差,一些较新的研究无法收录;(2)数据库中的信息不完整,了解某分子或者某信号通路,不能单单依靠数据库信息。因而,依然需要阅读大量领域相关文献。虽然各大高校都购买了一些出版社的下载权限,但是依然存在需要阅读的文献无法下载的情况。此外,高校购买的下载权限一般必须在校园

网内才提供下载,互联网无法下载,影响研究生查阅文献的效率。近年来,SCI-HUB把出版社搞得焦头烂额,SCIENCE杂志甚至发文称SCI-HUB让其“爱恨交加”^[12]。就实用性而言,只需要输入相应文献标题或者页卷号,就可以得到需要的文献全文,对于提高研究生科研效率是毋庸置疑的。

1.3 微信公众号科研启蒙 对于刚进入实验室的低年级研究生,会面临许多实际存在的问题,而导师并不可能时刻协助解决,这就需要研究生自行寻找解决问题的方式,也是培养研究生科研素养的重要训练。近年来,随着中国政府科研投入的日益增长,进行科学研究工作人群数量提高,衍生出以科研技能培训的一个群体。特别是中国微信时代下,借助于微信平台,许多科研启蒙平台应运而生^[13],包括“小张聊科研”“实验万事屋”等,以一种诙谐幽默,通俗易懂的方式,免费介绍文献阅读,课题申报,数据库分析,常用软件使用等科研工作者时常涉及的基本内容。这些公众号把复杂的科学研究形象化,简单化,特别适合研究生理解学习科研工作的基本思路和逻辑线,在研究生起步阶段帮助很大。此外,在免费的基础上,一些公众号也提供如信号通路图模板,生物信息学软件教程等付费服务,可以满足研究生在文章撰写阶段的需求。

2 信息化社会为病理研究生科研能力培养带来新的挑战

海量的数据,文献,资讯给肿瘤病理学研究带来了巨大的冲击,如何在有限的时间内,高效的获取自己需要的信息成为新的挑战^[14]。以往的培养模式,更多的强调理论水平与实验操作能力,而忽略了计算机分析能力的培养,在研究生的课程中也鲜有涉及^[15]。而计算机水平的高低,极大的决定了其获取理论知识,掌握实验技能,分析生信数据的效率。因而,应该通过“三位一体”的新型培养方式,把理论和实践水平作为基础和支撑,计算机应用能力作为保障,逻辑思辨能力作为核心,才能培养出适应新时代需要的肿瘤病理学人才。

2.1 提高研究生利用计算机处理海量数据的能力 下载的TCGA肿瘤数据库的转录组学数据,如何才能筛选到有关键分子;送检样本的蛋白组学数据,如何才能分析出下游的信号通路;收集的肿瘤患者临床信息资料,如何需要核心的预后指标。这些都需要研究生能够熟练运用计算机进行数据的处理。除了基础的EXCEL, SPSS, Treeview等软件的使用,网上数据库也提供在线数据分析功能,如GSEA, DAVID都能进行信号通路分析。这些

大数据的分析研究,如同第三方验证,更好支撑我们的研究结论,是现今肿瘤研究不可或缺的一部分。因而培养研究生掌握利用计算机处理海量数据的能力,是提高研究生能力的重要方式。

2.2 强化研究生图表呈现能力 所有的科研数据,需要简洁明了的图表;论文的科学假设,需要直观易懂的模式图;论文的编辑,需要标准规范的方式。如何把数据分析的结果,以一种直接美观的方式呈现,是一门新的学问。我们的课程设置和研究生培养要求中,对文献检索,课题设计,实验操作等内容提出了高质量的要求,却忽略了科研数据作图,科学假设模式图的内容。无法熟练地操作基本的作图和统计软件,更遑论用直观的图表展现大样本数据。纵观高质量的SCI论文杂志,图表和模式图无一不是美观精致。研究生在进行课题研究的时候,应该学习以高质量论文的图表方式呈现研究结果,平时多留心互联网的模式图模板,利用PPT或者PS进行元件拆分,在绘制模式图时才能加以利用。

2.3 始终强调研究生逻辑思辨能力培养 研究生的模块化教学中,文献检索和阅读,是培养研究生自主学习研究领域最新成果的能力;实验操作,是培养研究生动手能力;切片阅读,是培养研究生病理诊断能力^[16]。高度信息化的社会,慕课,爱课程等平台把系统化的大学课程带向千家万户,腾讯课堂^[17],网易云课堂也有面向研究生培训课程,包括SCI论文写作,实验操作,生物信息学分析等。知识的获取变得便捷无比,使得即使在读本科生,也能完成科研工作。但是科研工作不是单单读文献,做实验,统数据,更为核心的是,通过完整的课题研究,培养学生的逻辑思辨能力。并不是所有课题都像计划的一样无比顺利,真实的研究课题,会出现许多与假说相悖的实验结果,在排除技术问题的前提下,如何合理的解释,修正课题的研究方向,利用所在平台的资源继续推进研究工作,这才是研究生在三年乃至六年的学习中应该重点掌握的内容。方法会过时,热点会冷却,只有思路才是维持科研工作锐意进取,推陈出新的不竭动力。所以,作为导师,不应该为了课题进展速度,就全部包办,让研究生沦为实验的机器;而应该循循善诱,让研究生经历磨难的“阵痛”,才能收获能力的提升。

3 展望

大数据时代的来临,为肿瘤病理学研究生的培养带来了机遇,也带来了挑战。机遇,在于大数据时代的共享理念,让知识的传播变得方便,数据获取更为便捷;挑战,在于研究生的培养目标不断提

升,导师的能力要求不断加强。利用互联网丰富的资源,采用“三位一体”的培养模式,提高理论与实践水平,强化计算机应用技能,坚持逻辑思辨能力培养,才能训练出满足新时代要求的肿瘤病理学研究生。

利益冲突 作者申明不存在利益冲突

作者贡献声明 罗韬:提出设计思路、实施项目及撰写论文;姚小红:修订思路及实施项目;姚小红:实施项目及审校论文;时雨、阎晓初、卞修武:指导项目实施

参考文献

- [1] 马向涛.古病理学与肿瘤学发展史[J].肿瘤防治研究,2015,42(3):215-218.
- [2] 阮志华.从精准医学看肿瘤学教育[J].安徽医药,2017,21(3):584-586.
- [3] 郭毅可,杨钰.精确医学与大数据[J].上海大学学报(自然科学版),2016,22(1):17-27.
- [4] 张繁霜,应建明,鲁海珍等.肿瘤分子病理诊断中大数据的应用研究[J].中华病理学杂志,2018,47(7):562-564.
- [5] 师岩,张春晶,吴丽娜,等.“互联网+”时代的智慧教育在医学教育中的实践探索[J].中华医学教育杂志,2018,38(4):566-569,574.
- [6] 叶亮,李慧娟,张方,等.基于Oncomine数据库荟萃分析KIF23在非小细胞肺癌中的表达和意义[J].中国肺癌杂志,2017,20(12):822-826.
- [7] 张卓,江帅,李睿江,等.面向肿瘤精准医学的综合数据资源TCGA及其相关在线分析工具推荐[J].中华医学图书情报杂志,2018,27(3):5-9.
- [8] STELZER G, ROSEN N, PLASCHKES I, et al. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analysis [J]. Current Protocols in Bioinformatic, 2016, 54: 1.30.1 - 1.30.33. DOI: 10.1002/cpbi.5.
- [9] UHLÉN M, FAGERBERG L, HALLSTRÖM BM, et al. Proteomics. Tissue-based map of the human proteome [J]. Science, 2015, 347(6220): 1260419.
- [10] UHLEN M, ZHANG C, LEE S, et al. A pathology atlas of the human cancer transcriptome [J]. Science, 2017, 357(6352) DOI: 10.1126/science.aan2507.
- [11] BOWMAN RL, WANG Q, CARRO A, et al. GlioVis data portal for visualization and analysis of brain tumor expression datasets [J]. Neuro Oncol, 2017, 19(1): 139-141.
- [12] MCNUTT M. My love-hate of Sci-Hub [J]. Science, 2016, 352(6285):497.
- [13] 韩丽君,王国薇,潘思文等.病理学实验教学中引入微信公众平台的应用研究[J].中华医学教育探索杂志,2017,16(6):560-563.
- [14] 谢芳.基于互联网的大数据挖掘关键技术分析[J].信息记录材料,2018,19(10):95-96.
- [15] 梁洁,胡萍.基于医学生创新素质培养的计算机语言类课程改革探讨[J].中国卫生事业管理,2018,35(7):525-527.
- [16] 时雨,姚小红,平轶芳等.基于专项训练提高病理研究生科研新能力的策略[J].中华医学教育探索杂志,2017,16(8):792-795.
- [17] 尹文君.腾讯课堂网络平台在教学中的应用[J].科技展望,2017,27(8):204.

(收稿日期:2018-09-29,修回日期:2019-10-16)