

· 教育培训 ·

浅谈临床生物信息学在临床医学教育中的作用及初步实践

张皓旻, 杨波, 郭斌, 于睿莉, 迟小华, 陈红飞, 席义博, 陈锡勤, 贺培凤, 卢学春

[摘要] 生物信息学是一门整合计算机科学及生命科学的交叉学科,已有30余年历史,但仍方兴未艾。21世纪以来,生物医学已逐渐进入大数据时代,海量的以测序为代表的基因组、转录组、蛋白质组数据将为深刻揭示人类疾病发生机制及寻找治疗药物提供了前所未有的契机。目前,在临床医学研究及教学中,生物信息学方法及应用尚未得到足够重视。作者结合临床生物信息学发展现状及未来趋势,探讨了临床生物信息学学科在临床医学教学及转化研究中的作用。

[关键词] 临床生物信息学;医学教育;基因组学;数据库

[中图分类号] R-42 **[文献标志码]** A **[文章编号]** 2095-3097(2018)03-0175-04

doi: 10.3969/j.issn.2095-3097.2018.03.013

The role and practice of clinical bioinformatics in medical education

ZHANG Haomin^{1,2}, YANG Bo², GUO Bin^{1,2}, YU Ruili³, CHI Xiaohua⁴,
CHEN Hongfei², XI Yibo^{1,2}, CHEN Ximeng^{1,2}, HE Peifeng¹, LU Xuechun²

(1. Management School, Shanxi Medical University, Taiyuan Shanxi 030002, China; 2. Department of Hematology, Nanlou Division, National Clinical Research Center for Geriatric Diseases, Chinese PLA General Hospital, Beijing 100853, China; 3. Department of Allergy, Beijing Shijitan Hospital Affiliated to Capital Medical University, Beijing 100038, China; 4. Department of Pharmacy, Chinese PLA Rocket Force General Hospital, Beijing 100800, China)

[Abstract] Bioinformatics is an interdisciplinary subject integrating computer science and life science. It has been more than 30 years old, but it is still in the ascendant. Since the 21st century, biomedicine has gradually entered the era of big data, massive sequencing as a representative of the genome, proteomic data will reveal human disease mechanism and provide an opportunity for hitherto unknown drugs. At present, the methods and applications of bioinformatics have not been paid enough attention in clinical research and teaching. Based on the current status and future trends of clinical bioinformatics, this paper discusses the role of clinical bioinformatics in clinical medicine teaching and translational research.

[Key words] Clinical bioinformatics; Medical education; Genomic; Database

生物信息学是20世纪80年代兴起的一门新的应用生物学学科,在过去的30年里有了长足的发展,在生物学、基础医学、农学和法医学等领域都有了深入和广泛的应用,但在指导临床医学教学和临床转化研究方面应用相对不足。

临床生物信息学与经典生物信息学不同,侧重

于临床密切相关的生物信息学内容,包括疾病的诊断、治疗、预后评估等组学数据库、分析平台具体应用,是一门生物信息学与临床医学的交叉学科,是大数据研究的一个新分支。一言以蔽之,临床生物信息学是将组学大数据作为研究对象,结合临床具体问题,经数据挖掘,获得能够指导临床诊断、治疗及预后评估的知识。当前,随着计算机科学的进步,出现包括数据库分析、机器学习以及人工智能等一系列新技术,极大的丰富了临床生物信息学方法学。我们自2003年以来,利用临床生物信息学方法进行了一系列临床转化研究,并在国内多所医学院校主讲了有关临床生物信息学方法及应用的课程,在这些实践基础上,论述临床生物信息学在临床医学教育中的作用。

[基金项目] 国家自然科学基金面上项目(81273597,81302801);解放军总医院临床科研扶持项目(2016FC-ZHCG-1004);北京市科委首都临床特色重点课题(Z161100000516006)

[作者单位] 030002 山西太原,山西医科大学管理学院(张皓旻,郭斌,席义博,陈锡勤,贺培凤);100853 北京,中国人民解放军总医院南楼血液科,国家老年疾病临床医学研究中心(张皓旻,杨波,郭斌,陈红飞,席义博,陈锡勤,卢学春);100038 北京,首都医科大学附属北京世纪坛医院变态反应科(于睿莉);100800 北京,中国人民解放军火箭军总医院药剂科(迟小华)

[通讯作者] 卢学春, E-mail: luxuechun@126.com; 贺培凤, E-mail: hepeifeng2006@126.com

1 临床生物信息学的教育现状和迫切需求

随着基础研究的不断深入,基因组学、转录组学、表观基因组学和蛋白质组学的数量不断增多,全球范围内每天产生海量的DNA、RNA、蛋白质以及由此产生的二级数据库等生物信息数据。临床医生如何利用这些基础研究的数据,解决临床实践过程中诊断、治疗和预后评估的疑难问题,为患者服务,这是每一个基础研究和临床工作者所面临的难题。高效实用且容易掌握的生物信息学方法,是突破这些数据快速转化为临床信息的桥梁。临床生物信息学就是在此背景条件下应运而生的一门新的学科^[1]。

在当今转化医学逐渐引起学者们广泛关注,国外在这方面研究起步较早,方法相对成熟,而国内则刚刚起步,研究方法欠缺,且多数学者对此并不了解。为此,尽早开展有体系的临床生物信息学教学,以解决目前临床医生科研方法单一、临床研究缺乏创新性以及实用性等现实问题迫在眉睫。

2 高效临床生物信息学数据库

在经典的生物信息学教学当中,对计算机技术的掌握往往要求很高,一个优秀的生物信息学学者应当掌握大量的生物信息学数据库,能够熟练应用多种计算机编程语言进行数据的建库、清洗、处理等工作。但是,对医学生或临床医生而言,在学习高深的生物医学知识、处理繁杂的临床工作的同时,熟练掌握计算机技术显然是很难的。因此,对于生物医学工作者更重要的是掌握大量的生物信息学数据库及在线分析平台等工具,并能够利用这些工具对自己的医学研究、临床科研甚至于临床疑难病症的诊治进行指导,尤为重要。

临床生物信息学数据库根据其数据源可以分为一次数据库、二次数据库和在线分析工具。其中,存储有DNA序列、蛋白质序列、结构等信息的数据库构成了一次数据库。二次数据库是在一次数据库的基础上经实验验证及数据分析得到的,其中部分二次数据库还具有特有的分析工具。在线分析工具是以一次数据库和二次数据库为背景参照,运用自主设计的算法来提供在线分析功能。一般而言,一次数据库具有种类多样、数据更新和增长快、结构复杂且层次深、高度计算机化和网络化等特点;二次数据库则针对性强,大多提供检索及开源下载,使用方便;在线分析工具除具有一定的在线分析功能外,还具有原始数据检索和可视化功能,同时可供用户下载分析结果^[2]。

2.1 一次数据库 GenBank 数据库(<http://www.ncbi.nlm.nih.gov>)、欧洲分子生物学实验室(The European

Molecular Biology Laboratory, EMBL) 核酸库(<http://www.ebi.ac.uk/ena/>)和日本的DNA数据库(DNA Data Bank of Japan, DDBJ) (<http://www.ddbj.nig.ac.jp/>)是著名的核酸序列数据库。GenBank 数据库是美国国立生物技术信息中心建立的,包含所有已知的DNA和蛋白质序列,以及与其相关文献和注释。EMBL 核酸库是由欧洲生物信息学研究所维护的。DDBJ 数据库可与 GenBank 数据库、EMBL 核酸库交互,也是一个非常全面的DNA序列数据库。

蛋白质序列数据库有 Swiss-Prot (<http://www.ebi.ac.uk/uniprot>)、PIR-PSD (<http://pir.georgetown.edu/>)和蛋白质数据库(Protein Data Bank, PDB) (<http://www.rcsb.org/>)等。Swiss-Prot 是欧洲最大的蛋白质序列数据库,它的序列准确率高,注释完善,但与其他数据库相比数据量较小,现在已整合进 UniProt 数据库中。PIR-PSD 是国际上最大且有注释信息的开源蛋白质序列数据库。PDB 是国际上唯一的生物大分子结构数据库,其数据来源于X光晶体衍射和核磁共振,经过整理和确认后存档而成。

2.2 二次数据库 二次数据库数量很多,如基因表达综合数据库是目前完全公开的最大的高通量基因表达数据库;人类基因综合数据库 GeneCards 提供了全部基因、蛋白质在转录、遗传和功能上所有已知信息。GeneCards 中的信息功能包括某一基因与疾病的关系,基因多态性、表达、功能、分布,蛋白质与蛋白质相互作用等,且提供各数据库及分析工具链接,操作简单;miRBase 数据库是一个提供包括 miRNA 序列数据、注释、调控基因等信息的数据库;蛋白质数据库,它包含提供全部24 000种人类蛋白质的组织和细胞分布信息。

此外,尚有一些关于化合物或代谢物数据库,如代谢物、药物和毒性分子数据库 fragmentstore (http://bioinf-applied.charite.de/fragment_store);在实验基础上建立的,提供与癌症有关的蛋白和化合物相互作用的 Cancer Resource 数据库 (<http://data-analysis.charite.de/care/>)等。

2.3 在线分析工具 在所有在线分析数据库当中,DAVID(the Database for Annotation, Visualization and Integrated Discovery, <http://david.ncifcrf.gov/>)是临床生物信息学最常用的聚类分析数据库,它整合了生物学数据和分析工具,帮助用户从中获取有用的生物学信息。该工具在2003年发布,目前版本是v6.8。与其他分析工具相似,都是利用统计方法,在数千个基因关联的注释中,找出关联度最高的生物学注释,最主要是功能注释和信息链接。

Reactome (<https://reactome.org/>)是一个开源

的、开放的、手动策划和同行评议的途径数据库,是一个直观的生物信息学工具。用于可视化、解释和分析通路知识,以支持基础和临床研究、基因组分析、建模、系统生物学和教育。用户可从中直接检索生物学通路,查看生物学进程中发生的具体事件,也可以对用户输入的基因列表进行聚类,查看输入基因在各通路中的具体作用。

STRING 数据库(<https://string-db.org/>)是一个检索已知蛋白质相互作用的工具。蛋白质相互作用既包括蛋白质之间理化性的直接相互作用,也包括蛋白质之间功能性的间接相互作用。它不仅有实验结论,还有对 PubMed 进行文本挖掘得出的结论,同时综合其他数据库信息,利用生物信息学的方法进行分析。该工具会对不同来源的结果基于一定的评分方式给出不同的权值,最终根据其特有的评分方式给出一个最终打分。

3 利用临床生物信息学方法指导基础研究与临床实践

随着各种类型和层次大数据的不断出现,基因组学和二代测序技术的发展,组学大数据以指数级增长。各种类型、功能各异的数据库层出不穷。对于一个医学生、医学科研和临床工作者来说,如何选择合适的数据库和方法指导医学研究,如何从数据库中繁杂的信息中提取出有价值的知识,是每一个医学生、医学科研和临床工作者应当具备的能力。

3.1 利用生物信息学数据库对序列进行检索 对于许多未知生物学功能的新序列,研究人员从数据库中检索出与新序列同源的已知序列,根据已知同源序列推测出新序列生物学功能。而临床工作者则需要知道该序列在不同组织,不同状态下的表达情况。这也是二者面对同一序列,但研究目的截然不同的主要原因。前者属于经典生物信息学的研究内容,而后者则是临床生物信息学的研究方向。在分子生物学中,DNA 或蛋白质的相似性是多方面的,可能是结构相似,也可能是功能相似。所以在进行序列相似性分析时,希望从一些普遍规律中推测出新序列的结构或功能,从而发现生物分子的新内涵。这种方法在大多数情况下是成功的,当然也有例外,因此这种基于相似性的分析方法仍然需要不断改进^[3]。

3.2 利用生物信息学数据库进行转化研究的浅层次数据挖掘 数据挖掘就是从大量、不完全、有噪声、模糊的数据中,提取隐含在其中未知但又有效、新颖、有潜在价值、并且最终转化为信息和知识的过程。目前,随着多种组学的出现,数据量不断增加,数据库日益增多,组学大数据分析方法的出现为序

列分析、基因表达分析、蛋白质结构预测以及新药研发等领域的数据挖掘提供了广阔的应用前景。利用数据挖掘技术,将帮助人们更好地认识组学大数据所蕴涵的生命信息,探究生命的奥秘,并将此应用于临床,为更多患者解除病痛的折磨,为医学的进步注入新的活力。

对此,本研究团队早在 2003 年就开始了探索。早期,我们通过应用简单的临床生物信息学方法对基因组及转录组学数据进行数据挖掘,获得了一系列喜人的原创成果,且多数成果已成功应用于临床,为后续临床研究提供宝贵经验。例如,筛选并分析了急性淋巴细胞白血病复发相关基因^[4];预测并验证了去甲基化和组蛋白去乙酰化转移酶抑制剂对肿瘤细胞的影响^[5];观察氨磷汀联合重组人红细胞生成素对高龄骨髓增生异常综合征的疗效^[6];结合中医“辨证施治”思想利用临床生物信息学分析对免疫性血小板减少性紫癜进行分型施治^[7]。

3.3 利用生物信息学数据库进行转化研究的深层次知识发现 随着组学数据的日渐增多,简单的临床生物信息学挖掘方法已不能适应大数据量、大运算量、多层次组学数据的深层次分析。随着计算机技术的不断进步,国内外先后有研究团队将机器学习、人工智能等方法应用于临床生物信息学研究,实现对多组学数据的更深层次分析,以期指导医学研究、临床诊治和新药研发。机器学习的目的是实现在海量数据中由计算机自动的对潜在知识进行挖掘,因此,海量数据是实现机器学习以及确保所获取知识准确性的必要前提。

为此,本研究团队建立了疾病多组学数据库、药物多组学数据库以及医学文献数据库,以此为基础利用机器学习等大数据分析技术自主建立了“疾病-药物多组学大数据临床生物信息学平台”,进行血液病及肿瘤治疗药物和方案的系列研发,并成功应用于临床^[8]。利用这种方法,我们最先对依硫磷酸联合方案治疗骨髓增生异常综合征进行优化^[9-10]。随后,陆续在国际上,率先报道 6 种创新治疗技术体系:①反复多疗程自体免疫细胞治疗技术体系^[11];②超低剂量表观遗传药物联合免疫治疗技术体系^[12-16];③含祛脂向分化药物的再生障碍性贫血联合治疗技术体系^[17-20];④靶向活化性免疫细胞治疗技术体系^[21-23];⑤靶向细胞周期素 D1 阳性肿瘤的治疗技术体系^[24];⑥泛细胞保护剂联合造血生长因子方案治疗骨髓增生异常综合征及免疫性血小板减少症的技术体系^[25-26]。此外,本研究团队在多所医学院校进行临床生物信息学教学中发现:①各类疾病指南过多,限制医学生创新思维;②生物医学数据量大,但被利用转

化为临床可用的信息少,不利于大数据时代下医学教育的发展。正如古人所说,授人以鱼不如授人以渔,作者认为,当务之急在临床医学教育中开展临床生物信息学教学及方法研究。

4 结语

随着人类基因组计划的完成和二代测序技术的发展,运用机器学习等前沿技术针对爆炸式增长的组学大数据进行深度挖掘成为可能。面对传统治疗方案难以解决临床疑难重症的严峻现实,医学本科毕业生有必要开展临床生物信息学教学,有体系地学习临床生物信息学方法,了解并熟练掌握临床生物信息学常用的数据库。使得每一位临床医学研究生、医学工作者对现有医学研究和临床工作中难以解决的医学问题进行合理分析并能有效解决。最终使临床生物信息学能满足现代医疗和医学研究发展的需要,促进临床转化,为现代医学的发展提供新途径。

【参考文献】

[1] Trent RJA.临床生物信息学[M].卢学春,杨波,张峰,译.北京:军事医学科学出版社,2014.

[2] 刘银凤,张雷.生物信息学数据库在医学研究中的应用[J].中国病原生物学杂志,2014,9(10):附页1-2,封三.

[3] 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.

[4] 卢学春,楼方定,徐周敏,等.急性淋巴细胞白血病复发相关基因的筛选及生物信息学分析[J].军医进修学院学报,2004,25(2):81-83.

[5] 卢学春,楼方定,徐周敏,等.去甲基化和组蛋白去乙酰化转移酶抑制剂对 K562 细胞增殖和肿瘤相关基因表达的影响[J].中国实验血液学杂志,2004,12(1):44-47.

[6] 卢学春,朱宏丽,姚善谦,等.氨磷汀联合重组人红细胞生成素治疗高龄骨髓增生异常综合征近期疗效观察[J].中国实验血液学杂志,2005,13(3):440-442.

[7] 卢学春,朱宏丽,姚善谦.免疫性血小板减少性紫癜分型施治的基础与临床研究进展[J].中国实验血液学杂志,2008,16(5):1232-1236.

[8] 中国癌症基金会《中国肿瘤临床年鉴》编辑委员会.2016 中国肿瘤临床年鉴[J].北京:中国协和医科大学出版社,2017,578-579.

[9] 卢学春,杨波,朱宏丽,等.生物信息学方法优化依硫磷酸联合方案治疗骨髓增生异常综合征的应用研究[J].中华医学杂志,2009,89(26):1834-1837.

[10] 杨波,蔡力力,迟小华,等.依硫磷酸调控人类基因表达谱的预测及生物信息学分析[J].中国实验血液学杂志,2011,19(3):711-716.

[11] Yang B, Lu XC, Yu RL, et al. Repeated transfusions of autologous cytokine-induced killer cells for treatment of haematological malignancies in elderly patients: a pilot clinical trial[J]. Hematol Oncol, 2012, 30(3): 115-122.

[12] 卢学春,杨波,朱宏丽,等.自体细胞因子诱导的杀伤细胞联合 IL-2 治疗老年人血液系统恶性肿瘤的临床经验探讨[J].解放军医学杂志,2010,35(10):1270-1272.

[13] Lu XC, Yang B, Yu RL, et al. Clinical study of autologous cytokine-induced killer cells for the treatment of elderly patients with diffuse large B-cell lymphoma[J]. Cell Biochem Biophys, 2012, 62(1): 257-265.

[14] 蔡力力,杨波,卢学春,等.免疫功能检测对评估细胞因子诱导的自体杀伤细胞治疗老年血液肿瘤疗效的研究[J].中国实验血液学杂志,2010,18(5):1250-1255.

[15] Yang B, Wang HT, Cai LL, et al. Successful management of acute myeloid leukemia transformed from myelodysplastic syndromes in an elderly patient aged over 80 years old by ultralow dose decitabine combined with amifostine and autologous CIK cells[J]. Ann Hematol, 2013, 93(7): 1233-1235.

[16] Yang B, Yu R, Cai L, et al. A comparison of therapeutic dosages of decitabine in treating myelodysplastic syndrome: a meta-analysis[J]. Ann Hematol, 2017, 96(11): 1811-1823.

[17] 卢学春,迟小华,杨波,等.重型再生障碍性贫血发病相关 T 淋巴细胞基因表达谱的生物信息学分析及作为药物筛选新方法的探索[J].中国实验血液学杂志,2010,18(2):416-420.

[18] 卢学春,杨波,迟小华,等.含盐酸二甲双胍联合方案治疗再生障碍性贫血的短期疗效观察[J].解放军医学杂志,2012,37(3):229-233.

[19] 卢学春,杨波,迟小华,等.再生障碍性贫血病因学的新探索:异常免疫诱导骨髓间充质干细胞的过度脂肪化(英文)[J].解放军医学杂志,2014,39(3):173-179.

[20] 卢学春,迟小华.含盐酸二甲双胍和维生素 b12 的组合药物:中国, CN 101716182 A[P]. 2010-06-02.

[21] 蔡力力,杨洋,杨波,等.含胸腺肽增强免疫的自体 CIK 细胞输注联合小剂量 IL-2 方案治疗老年人 B-CLL 的近期疗效观察[J].中国实验血液学杂志,2012,20(3):564-570.

[22] Yu R, Yang B, Chi X, et al. Efficacy of cytokine-induced killer cell infusion as an adjuvant immunotherapy for hepatocellular carcinoma: a systematic review and meta-analysis[J]. Drug Des Devel Ther, 2017, 11: 851-864.

[23] 杨洋,杨波,脱帅,等.含胸腺肽免疫增强的自体 CIK 细胞联合 IL-2 方案治疗高龄弥漫大 B 细胞淋巴瘤[J].军医进修学院学报,2012,33(5):441-443.

[24] 杨波,卢学春,迟小华.氨磷汀在细胞周期素 D1 高表达类型肿瘤中的应用:中国,201510142245.2[P]. 2015-10-14.

[25] 中国癌症基金会《中国肿瘤临床年鉴》编辑委员会.2015 中国肿瘤临床年鉴[J].北京:中国协和医科大学出版社,2016,200-208,669-670.

[26] 于睿莉,杨波,蔡力力,等.老年骨髓增生异常综合征的临床病理特征及氨磷汀联合造血生长因子方案的长期疗效观察[J].中国药物应用与监测,2017,14(2):75-80.