

基于医药大数据建立精准药学临床评估体系的思考

王伟¹, 祁永飞¹, 包晗¹, 邱鹏程¹, 张鑫磊¹, 王玉琨^{1,2*} (1. 空军军医大学药学院药事管理与药物信息学教研室, 西安 710032; 2. 空军军医大学药学院, 西安 710032)

摘要 目的: 探讨医药大数据服务于精准药学的现有问题和对策。方法: 通过文献资料, 分析精准药学临床应用所需的大数据及其发展现状, 从药物基因组学研究和个体基因组学研究的视角, 探讨目前个性化用药研究的局限性。结果与结论: 现有的药物基因组学研究成果远远不能满足临床需求, 而基因组学研究成果过于碎片化, 难以应用于临床。整合临床电子健康记录 (EHR) 的基因组研究, 能反映药物研发和应用的实效; 建立依据EHR和组学数据的精准药学临床评估体系, 将大大提高精准药物研发和个性化用药的循证性和准确性。

关键词: 精准药学; 大数据; 个性化用药; 药物基因组; 临床评估

中图分类号: R95; TP 39 文献标识码: A 文章编号: 1002-7777(2017)12-1478-05

doi:10.16153/j.1002-7777.2017.12.018

Establishment of the Clinical Evaluation System of Precision Pharmacy Based on Medical Big Data

Wang Wei¹, Qi Yongfei¹, Bao Han¹, Qiu Pengcheng¹, Zhang Xinlei¹, Wang Yukun^{1,2*} (1. Department of Pharmaceutical Administration and Informatics, School of Pharmacy, Air Force Military Medical University, Xi'an 710032, China; 2. School of Pharmacy, Air Force Military Medical University, Xi'an 710032, China)

Abstract Objective: To explore the existing problems and countermeasures of medical large data service in precision pharmacy. **Methods:** Big data required in clinical application of precision pharmacy and development status at present were analyzed based on literature review and logical analysis. The limitations of current personalized medicine researches were explored from the perspectives of pharmacogenomics and individual genomics studies. **Results and Conclusion:** Current achievements of pharmacogenomics studies couldn't meet clinical requirement. The achievements of individual genomics studies were too fragmented to be used in clinic. The genomic studies combined with electron health records (EHR) could reflect the research and development of drugs and the effectiveness of medication to some extent. The establishment of the clinical evaluation system of precision pharmacy based on EHR and omics data would greatly promote the research and development of precision drug and the evidence-based accuracy of personalized medicine.

Keywords: precision pharmacy; big data; personalized medicine; pharmacogenomics; clinical evaluation

基金项目: 国家自然科学基金项目 (编号 81702733); 肿瘤生物学国家重点实验室项目 (编号 CBSKI2015Z14, CBSKI201713); 陕西省科技统筹创新工程 (2016KTCL03-09)

作者简介: 王伟, 博士, 讲师; E-mail: wangwei_fmму@163.com

通信作者: 王玉琨, 博士, 教授; E-mail: wangyk@fmму.edu.cn

精准医学 (precision medicine) 的概念自2012年被提出之后, 各国陆续启动了精准医学计划; 2016年, 我国正式将精准医学研究纳入“十三五”发展规划重点专项计划, 提出要加快发展精准医学新模式。精准医学的研究范畴涵盖了多角度、多层面, 包括基础医学、临床医学、健康管理、生物和医药大数据研究等等, 旨在通过整合个体遗传信息、环境影响及生活习惯等多维度信息, 为达到临床早期诊断、疾病干预和个体化治疗提供科学依据^[1]。精准药理学在实现精准医疗的过程中起着重要作用, 它包含了药物研发和临床用药两个方面^[2]。目前, 个体化用药是精准医学研究、精准药理学研究在临床应用的主要方式之一。然而, 由于疾病种类多、单一疾病相关药物种类多, 要做到精准医学、精准药理学提倡的精确用药和个体化用药, 离不开医药大数据的支持; 要评价精准用药是否真正精准、个体化用药是否确实提高了疗效, 也离不开医药大数据的支持。

当今是大数据的时代。基因组学(genomics)、蛋白质组学(proteomics)、代谢组学(metabonomics)、暴露组学(exposomics)等系列组学研究数据, 提供了多方位视角来观察患病个体的疾病进展状态和进程; 医院电子信息系统如HIS等, 提供了精准医学、精准药理学服务所必须的人群就医、用药信息; 计算机技术、疾病相关生物信息学和医学统计学的发展, 提供了对接个体组学大数据和群体医药大数据的研究手段和方法。应用大数据时代新的思维方式和研究方法, 使其在开展精准医学研究的同时开展评价研究成为可能; 使临床药理学服务更具有循证性和科学性成为可能; 使个体化治疗更加精准、安全、规范、经济成为可能。

1 药物基因组学是个体化用药的理论基础

近几十年的药物基因组学研究发现, 不同个体所携带的遗传信息差异包括基因型差异、单核苷酸多态性 (single nucleotide polymorphism, SNP)、基因拷贝数变异 (copy number variation, CNV) 等会导致药物转运、代谢以及药物靶点敏感性的显著差异, 这是个体化用药的理论基础。比如, 细胞色素酶 CYP450 家族蛋白 (cytochrome protein-450, CYP450) 的表达水平会显著影响药物代谢和药效^[3]。当前, 个体化用药的模式是以基

础和临床医学研究成果为依据, 结合病人基因组数据信息指导病人用药, 以达到减少副作用、提高治疗效果的目的。典型的例子是华法林的使用剂量需要参照细胞色素酶基因 *CYP2C9* 和维生素 K 环氧化物还原酶基因 *VKORC1* 的基因型, 通过个体化用药可以有效降低副作用, 让用药安全有效。因为: *CYP2C9*2*、*CYP2C9*3* 基因型的患者细胞色素酶 CYP2C9 活性较低, 使用高剂量华法林会导致出血危险; *VKORC1* 基因启动子区点突变 (rs9923231) 会引起该酶的表达量减少, 携带该突变的患者需要降低剂量使用^[4]。

肿瘤化疗由于疗效个体差异大、毒副作用严重、价格高昂, 个体化用药在肿瘤治疗中的作用更为显著。比如, 拓扑异构酶抑制剂类化疗药物伊立替康在肝脏中代谢, 经尿苷二磷酸葡萄糖醛酸转移酶 (UGT) 降低毒性, 而 *UGT1A1*28* 基因型患者该酶的活性较低, 因此美国食品药品监督管理局 (FDA) 提醒该基因型患者有造成消化道黏膜损伤和腹泻的高风险^[5]; 他莫昔芬是雌激素调节剂类化疗药物, 通过 CYP450 家族酶在体内代谢转变为活性产物而发挥抗肿瘤活性, 然而 *CYP2D6* 基因具有40多种突变, 大多数突变方式会降低他莫昔芬活性产物的活性, 因此根据个体遗传信息, 美国FDA将患者人群分为快、中、慢3种代谢型, 快代谢性人群使用他莫昔芬将获得更好的疗效^[5]; 又如, 5-氟尿嘧啶 (5-fluorouracil, 5-FU) 的代谢限速酶二氢嘧啶脱氢酶 (DPD) 基因 *DPYD* 多态性与其毒副作用密切相关, 研究发现, 不同人种的基因突变频率差异较大, 白种人携带的 *DPYD*2A* 突变会使药物在消化道和血液系统中产生毒副作用, 但该突变在东亚人群中的基因频率几乎为零^[6]。

由此可见, 当前一些药物的国外研究已有较为确凿的结论, 相关基因型及多态性的检测甚至已经被FDA推荐作为药物使用规范; 然而, 还有许多临床常用药物虽然研究报道较多但临床数据尚不够充分, 且国外临床依据充分的数据信息对中国人群是否适用? 是否中国人群特有的基因型/多态性尚未被发现? 这些问题昭示着尚需要一套评价系统, 回顾性地反映药物基因组学研究成果在临床应用中的实效, 通过回顾性评价研究才能体现个体化用药的精确性和循证性。

2 精准药学的广泛应用离不开个体组学大数据

随着基因组计划的完成,人类基因组信息已被多个数据库公布,如NCBI(National Center for Biotechnology Information, NCBI)、ENSEMBL(基因组数据库)、UCSC(University of California Santa Cruz, UCSC)等^[7],药物基因组学研究成果主要集中在美国卫生研究院建立的药物基因组学知识库PharmGKB(pharmacogenomics knowledge base)。PharmGKB涵盖了基因多态性注释、药物代谢及其他相关通路、药物剂量和使用指南等信息^[8]。到目前为止,已收录超过2.7万条基因注释、超过3700条化学药物信息、70条药物剂量使用说明、251条药物使用说明、112条药物通路。CPIC(clinical pharmacogenetics implementation consortium)数据库^[9]建立于2009年,在PharmGKB的基础上整合了PGRN(pharmacogenomics research network)的相关信息,公开发布了有实验和临床研究证据的药物基因组学临床应用指南,目前已发布33种药物指南、352条药物-基因关系、近400条等位基因涉及的临床用药信息。除了PharmGKB这样已经整合完成的数据库以外,还有很多药物基因组学研究正在不断地发现新基因型/多态性与药物的临床关系,这些研究高度依赖临床样本基因测序数据。随着基因芯片技术和DNA测序技术的发展,许多数据库将基因信息更细致地划分并根据收集的样本公布了新的数据。比如NCBI dbSNP数据库收录了基因组SNP位点信息;NCBI Clinvar数据库^[10]注释了部分基因突变与表型的关系;OMIM数据库^[11]收录了基因型与其对应疾病表型的注释,目前收录超过1.5万条基因注释、6000余条疾病表型;HGMD数据库^[12]注释了人类基因突变及与相应疾病的对应关系,目前收录了5600余种基因超过14万种突变;肿瘤细胞基因突变数据库COSMIC^[13]收录了肿瘤的体细胞突变及突变频率;TCGA数据库^[14]收录了34种肿瘤样本的基因组数据,平均每种肿瘤样本量超过300例;CellMiner数据库^[15]收录了60种人类肿瘤细胞系的基因组数据,并提供了生物信息学分析数据工具,等等。

遗传数据库中的基因数据主要来源于基因芯片和DNA测序,近几年应用第二代测序(next-generation sequencing, NGS)技术获得遗传信息已

成为遗传数据库的主要数据来源。NGS可以高通量、低成本地同时进行大量的DNA测序,该技术已经不仅应用于基础和临床医学研究,而且广泛用于个体化治疗。2015年,美国FDA建立了“Precision FDA”平台,这个平台是一个基于NGS技术的在线基因组信息数据库,而且还是一个基因组信息分析平台,研究人员可以开发、测试、共享基于NGS技术的数据信息和生物信息学工具,并可以验证他人的研究成果,从而促进了基于基因组数据的精准医学形成金标准和分析规范。

3 群体医药大数据整合使精准药学--临床评估成为可能

当前的基因组学研究即时反映了个体遗传信息—主要是基因组突变和临床表型的相互关系。然而,若能体现群体的遗传信息与临床表型的关系,将会使基因组学研究发挥更大的价值。2011年,美国哈佛大学医学院学者提出整合临床电子健康记录(electronic health records, EHR)如病历信息等,将促进基因组学研究成果的应用和普及,而且未来有望实现临床信息指导下基因组学研究和应用的标准化^[16]。目前,世界范围内已经建立了一些整合临床数据、药物使用信息和基因组数据的数据库,如美国范德堡大学开展的eMERGE(the electronic medical records and genomics)项目以及建立的世界最大的临床-基因整合数据库BioVU^[17]。基于BioVU数据库的研究,以临床病例为基础报道了华法林、氯吡格雷、他克莫司等药物疗效及副作用与基因多态性的关系^[18]。这些研究把临床信息与基因组学直接对接,循证地评价了药物基因组学在临床应用的可行性和有效性,但是全世界相关的研究都还较为匮乏。

关于肿瘤病人精准用药的研究,以往主要集中于临床试验阶段。例如,346项I期药物临床试验表明,采用精准用药策略治疗肿瘤病人将使其生存期平均中位数从3个月提升至5.7个月^[19];法国一项名为SHIVA的II期临床试验报道,据实验结果和理论推导具有靶向抗肿瘤效果,但未经临床试验批准的部分靶向抗肿瘤药物,则不能显著提升肿瘤患者无进展生存期的中位数^[20]。这些研究表明,目前靶向药物从研发到应用还有一定的盲目性,若建立一套评估体系,则会大大提高研发的效率和用药准确性。2017年,一项基于临床-基因组数

据库 (clinico-genomic database) 的肿瘤个体化用药研究在美国临床肿瘤学会年会 (American Society of Clinical Oncology, ASCO) 上发表^[21], 该研究利用临床-基因组数据库采集了1619例非小细胞肺癌病人的临床用药数据和基因组表达数据, 收集的病例平均年龄66岁、18%的患者高表达EGFR。研究发现, 应用美国国立综合癌症网络 (National Comprehensive Cancer Network, NCCN) 发布的非小细胞肺癌治疗指南中靶向药物治疗的病人, 平均中位生存期由28个月提高至42个月, 增长约50%。这些研究不仅表明精准用药对肿瘤患者治疗具有显著效果和重要意义, 而且是对靶向药物的再评估; 研究数据不仅启示我们在药物研发过程中及时评价的重要性, 而且启发我们评价系统可以使精准用药变得更加精准和个体化。

4 小结

精准药学的發展是推动精准医学普及的重要动力。当今的组学研究为认识人群遗传信息、了解个体用药的方式提供了便利的技术途径; 然而, 也使医学、药学研究愈发碎片化、远离临床医学。临床健康记录与个体数据的整合将搭起基础研究与临床应用的桥梁, 已经成为当前国际医学发展的大趋势。未来10年, 也许导致临床数据导向的基因组研究以及基础医学、药学研究发生重大变化的重要原因, 将是临床信息结合个体遗传、代谢等大数据新形式的整合数据的出现。可以预见, 新形式的整合数据将推动精准用药走向标准化、普及化; 组学研究将更重视临床样本的健康信息, 反过来也将进一步推动电子健康记录的建设 and 健全; 患者也许不仅能从精准药学指导的治疗中受益, 而且可能更多地参与到临床试验中以了解自己的遗传信息、尝试新的治疗方法, 从而最终使基因研究成果惠及大众, 推动医药事业蓬勃发展。

参考文献:

[1] Mirnezami R, Nicholson J, Darzi A. Preparing for Precision Medicine[J]. N Engl J Med, 2012, 366 (6): 489-491.
[2] 刘昌孝. 精准药学: 从转化医学到精准医学探讨新药发展[J]. 药物评价研究, 2016, (01): 1-18.
[3] Walsh JS, Miwa GT. Bioactivation of Drugs: Risk and Drug Design[J]. Annu Rev Pharmacol Toxicol, 2011,

51: 145-167.

- [4] Johnson JA, Caudle KE, Gong L, et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for Pharmacogenetics-Guided Warfarin Dosing: 2017 Update[J]. Clin Pharmacol Ther, 2017, 102 (3): 397-404.
[5] Swen JJ, Nijenhuis M, de Boer A, et al. Pharmacogenetics: from Bench to Byte--An Update of Guidelines[J]. Clin Pharmacol Ther, 2011, 89 (5): 662-673.
[6] Caudle KE, Thorn CF, Klein TE, et al. Clinical Pharmacogenetics Implementation Consortium Guidelines for Dihydropyrimidine Dehydrogenase Genotype and Fluoropyrimidine Dosing[J]. Clin Pharmacol Ther, 2013, 94 (6): 640-645.
[7] Hubbard T, Barker D, Birney E, et al. The Ensembl Genome Database Project[J]. Nucleic Acids Res, 2002, 30 (1): 38-41.
[8] Hewett M, Oliver DE, Rubin DL, et al. PharmGKB: the Pharmacogenetics Knowledge Base[J]. Nucleic Acids Res, 2002, 30 (1): 163-165.
[9] Relling MV, Klein TE. CPIC: Clinical Pharmacogenetics Implementation Consortium of the Pharmacogenomics Research Network[J]. Clin Pharmacol Ther, 2011, 89 (3): 464-467.
[10] Landrum MJ, Lee JM, Benson M, et al. ClinVar: Public Archive of Interpretations of Clinically Relevant Variants[J]. Nucleic Acids Res, 2016, 44 (D1): D862-D868.
[11] Hamosh A, Scott AF, Amberger JS, et al. Online Mendelian Inheritance in Man (OMIM), a Knowledgebase of Human Genes and Genetic Disorders[J]. Nucleic Acids Res, 2005, 33 (Database issue): D514-D517.
[12] Stenson PD, Ball EV, Mort M, et al. The Human Gene Mutation Database (HGMD) and its Exploitation in the Fields of Personalized Genomics and Molecular Evolution[J]. Curr Protoc Bioinformatics, 2012, Chapter 1: t1-t13.
[13] Forbes SA, Bindal N, Bamford S, et al. COSMIC: Mining Complete Cancer Genomes in the Catalogue of Somatic Mutations in Cancer[J]. Nucleic Acids Res, 2011, 39 (Database issue): D945-D950.
[14] Kim HS, Minna JD, White MA. GWAS Meets TCGA to

- Illuminate Mechanisms of Cancer Predisposition[J]. *Cell*, 2013, 152 (3) : 387-389.
- [15] Wang S, Gribkov M, Hazbun TR, et al. CellMiner Companion: An Interactive Web Application to Explore CellMiner NCI-60 data[J]. *Bioinformatics*, 2016, 32 (15) : 2399-2401.
- [16] Kohane IS. Using Electronic Health Records to Drive Discovery in Disease Genomics[J]. *Nat Rev Genet*, 2011, 12 (6) : 417-428.
- [17] Rasmussen-Torvik LJ, Stallings SC, Gordon AS, et al. Design and Anticipated Outcomes of the eMERGE-PGx Project: A Multicenter Pilot for Preemptive Pharmacogenomics in Electronic Health Record Systems[J]. *Clin Pharmacol Ther*, 2014, 96 (4) : 482-489.
- [18] 孙可欣, 詹思延, 胡永华. 医学大数据在药物基因组学领域中的应用与发展[J]. *药物流行病学杂志*, 2017, (01) : 68-73.
- [19] Schwaederle M, Zhao M, Lee JJ, et al. Association of Biomarker-Based Treatment Strategies With Response Rates and Progression-Free Survival in Refractory Malignant Neoplasms: A Meta-analysis[J]. *JAMA Oncol*, 2016, 2 (11) : 1452-1459.
- [20] Le Tourneau C, Delord J, Goncalves A, et al. Molecularly Targeted Therapy Based on Tumour Molecular Profiling versus Conventional Therapy for Advanced Cancer (SHIVA) : A Multicentre, Open-label, Proof-of-concept, Randomised, Controlled phase 2 trial[J]. *Lancet Oncol*, 2015, 16 (13) : 1324-1334.
- [21] Singal G, Miller PG, Agarwala V, et al. Development and Validation of A Real-world Clinicogenomic Database[Z]. ASCO Annual Meeting, 2017.

(收稿日期 2017年6月15日 编辑 王萍)