

· 综 述 ·

肿瘤生物信息学:系统临床医学的新手段

周洁白,王向东

[摘要] 肿瘤是临床上患者死亡最常见的原因,其诊断、治疗和预后情况取决于肿瘤的严重程度、病程、部位、对药物治疗的敏感性和耐药性、细胞的分化和起源以及对其病理过程了解程度的差异等多种因素。鉴于基因与蛋白质之间的联系和网络在肿瘤分子机制的研究中起重要作用,在肿瘤研究中引入了“系统临床医学”这个全新概念。肿瘤生物信息学根据疾病代谢、信号通路、信息交流以及增殖的特异性,将生物信息学方法应用于肿瘤研究,这种新兴的手段不仅是肿瘤系统临床医学的一个重要部分,而且是系统临床医学中进行肿瘤相关研究的核心手段,是改善肿瘤患者预后的一种最关键和最有效的方法。

[关键词] 肿瘤;生物信息学;系统临床医学

[中图分类号] R73;R-05

[文献标志码] A

[文章编号] 2095-3097(2012)01-0041-03

doi: 10.3969/j.issn.2095-3097.2012.01.011

肿瘤是临床上患者死亡最常见的原因,是发生在机体某个系统多个器官、某个器官多个系统,或者这两种情况共同存在的一种复杂的疾病。肿瘤的严重程度、病程、部位、对药物治疗的敏感性和耐药性、细胞的分化和起源以及对其病理过程了解程度的差异,决定了肿瘤的诊断、治疗和预后情况。越来越多证据表明基因与蛋白质之间的联系和网络在肿瘤分子机制的研究中起重要作用,因此把“系统临床医学”这个全新的概念引入肿瘤研究至关重要。系统临床医学整合了系统生物学、临床科学、谱学技术、生物信息学和计算机科学,改善疾病的诊断、治疗和预后情况。肿瘤生物信息学是肿瘤系统临床医学的一个重要部分,同样也是系统临床医学中进行肿瘤相关研究的核心手段。“肿瘤生物信息学主题系列”集合了 *BMC Bioinformatics*、*BMC Cancer*、*Genome Medicine* 和 *Journal of Clinical Bioinformatics* 等杂志的长处,关注肿瘤生物信息学在生物信息学方法、网络生物标志物和精准医学发展中的应用。“系列”聚焦肿瘤生物信息学和计算机系统生物学的新发展,探索其临床应用的潜力,以期改善肿瘤患者的预后。

1 方法学预期

肿瘤生物信息学是根据疾病代谢、信号通路、信息交流以及增殖的特异性,将生物信息学方法应用于肿瘤研究的一种手段。临床生物信息学,一门结合了临床信息学、生物信息学、医学信息学、信息技术、数学和谱学的新兴学科^[1],是解决临床上肿瘤

患者早期诊断、有效治疗和预后评估这些相关挑战关键因素的方法。有必要引入肿瘤生物信息学特异性方法学,或者引入全新、高级的生物信息学工具,来回答肿瘤相关的问题。例如,语义网技术的运用可以帮助分析从 Corvus 检索到的大量临床数据,建立定量的语义模型。Corvus 是一个数据库,为不同形式的谱学数据提供一个统一的平台,基于系统生物学知识以及 SPARQL (simple protocol and resource description framework query language) 节点的应用^[2]。语义模型,包含恶性黑色素瘤患者的基因组、转录物组和表观基因组数据,以及构建于转录因子结合信息的调节网络,用于研究不同分子阶段与其对于抗肿瘤治疗反应的相互作用。多元检验,一个阐明检验中由于样本准备而引入内源性误差并且检测共同作用因子的过程,用于帮助和指导临床医师了解基于 PAM50 的基因组预估在乳腺癌患者治疗计划中的应用,并且将不确定的信息用于一个可行的方案^[3]。

方法学、软件、计算机工具和数据库可用于研究肿瘤的分子机制、发现并证实肿瘤新的生物标志物、网络生物标志物和个体化药物,这些资源的适用性、特异性和整合性应予以关注。miRTrail 是一种整合工具,用于分析基因和 miRNA 间全面的交互联系,基于表达特征,从而得出更多关于失调的病理过程的有力证据和可靠结果。miRTrail 可能为研究人类疾病(包括肿瘤)进程中基因和 miRNA 间调节关系提供一种全新的手段,通过整合 20 000 个基因、近 1 000 个 miRNA 和大约 280 000 种普遍联系的信息而形成^[4]。探索可能的计算机方法,将基因和 miRNA 间的调节关系与临床表型联系起来,例如不同的肿瘤部位、肿瘤阶段、分化情况、患者症状,或者对治疗

[作者单位] 200032 上海,复旦大学附属中山医院呼吸科(周洁白),生物医学研究中心(王向东)

[通讯作者] 王向东, E-mail: xiangdong.wang@clintransmed.org

的反应中不同的基因间相互关系,可能对相关研究有所帮助。医学影像学应该成为肿瘤生物信息学应用中的一个重要因素,因为通过临床病理、超声、CT、MRI 以及 PET 等技术得到的影像,是肿瘤的“早期和准确”发现和诊断一个必要及重要的方法。对于医学影像中肿块和其他异常表现的形态学特征的生物信息学分析,开始于数学形态学对于目标特征的选择性获取,以及通过两种对照的修饰技术对于获取特征的强化^[5]。Haustein 和 Schumacher^[6]在“肿瘤生物信息学主题系列”中描述的计算机模型可以模拟肿瘤生长和转移形成的动态过程,从而在出现临床症状前发现可疑转移灶的形成。

肿瘤生物信息学专家帮助临床医师建立基因间或蛋白质间关系,以及与肿瘤的类型、密度或部位相关的一些机制的概念,可能是一个不相关的问题,或者说是一个未来的设想。von der Heyde 和 Beissbarth^[7]在“肿瘤生物信息学主题系列”中的一篇文章通过对表皮生长因子受体(epidermal growth factor receptor, EGFR)通路的分析,得出西妥昔单抗在头颈部肿瘤中耐药机制的最新进展。

2 生物标志物新策略

肿瘤生物信息学在生物标志物的发现和证实中起重要的作用,尤其在早期诊断相关的临床表型、判断疾病进展监测和治疗反应以及评估患者生活质量的改善方面。在肿瘤多种基于基因、蛋白质、多肽、化学、或者生理的变量中,生物标志物的研究从单个标志物到多个标志物,从表达到功能指标,从网络到动态网络。网络生物标志物是一类新的生物标志物,与蛋白质间关系有关,通过整合蛋白质注解、联系和信号通路的相关知识进行研究。网络生物标志物在疾病发展的不同阶段和时间点会有所改变,标志物的改变可以被检测和评估,即所谓的动态网络生物标志物,是一种新的研究方法。动态网络生物标志物可能与临床信息学有关,包括患者主诉、既往史、治疗、临床症状和体征、生化检查、影像学检查、病理以及其他一些检查手段^[8]。

系统临床医学在肿瘤生物标志物的研究发展中是推荐的一种新方法。系统临床医学整合了系统生物学、临床表型、高覆盖技术、生物信息学和计算机科学,以期改善疾病的诊断、治疗和预后情况。肿瘤生物标志物应该具有网络、动态、关联,以及对于疾病诊断、治疗和预后具有特异性等特征。认识临床信息学和生物信息学间的联系是研究发展疾病新的诊断和治疗手段的第一步,也是关键的一步。这种研究方法在肾移植后急性排斥反应和肺脏疾病的研

究中已经有所涉及^[9-10]。简而言之,临床研究中,按照清晰和严格的入组标准将研究对象入组,进行样本搜集,将临床描述转化为一套完整的临床生物信息学信息。通过生物信息学和系统生物学,可以分析样本的基因和(或)蛋白质特征,从而探索基因和(或)蛋白质间动态网络和相互联系。

通过计算机方法筛选出的具有疾病特异性的基因和(或)蛋白质网络和动态网络与各个临床表型分别相关,可以证实并完善疾病特异性的生物标志物。然而,在系统临床医学的应用中面临许多挑战需要克服。例如将临床描述信息转化成临床信息学信息的最优系统,以疾病严重程度、病程、部位、对治疗的敏感性和疾病进展为目标的生物信息学分析,或者是将来自临床的所有信息和高覆盖数据整合从而得出精确结论的计算机方法。除了寻找基因与蛋白质表达的变化和重要性,探索分子网络之间、分子网络和临床表型之间,以及基因和(或)蛋白质相互关系之间的变化和重要性也是一个挑战。Cun 和 Fröhlich^[11]在“肿瘤生物信息学主题系列”中报告,在一项将乳腺癌患者分级的研究中,结合蛋白质网络和关联的数据,提高了判断基因特点的能力,R 加权回归特征消去和平均通路表达是那些被检验的方法中发现可识别特点最有效的方法,这一发现证实了以上观点。

3 精准医学的检测和预期

系统肿瘤医学是认识并了解预言性、预防性、个体化和分享性(P4)医学的新手段^[12-15]。Tian 等^[15]最近提出,应用高覆盖技术将得出与患者相关的大量实体数据,包括疾病关联器官的细胞中一个或多个与疾病相关的分子网络。通过疾病相关的分子网络可以发现早期的异常信号,最终实现肿瘤的 P4 医学。肿瘤临床生物信息学是实现系统临床医学的一个重要的方法,通过将临床检查结果与体征和人类肿瘤组织来源的生物信息学信息结合起来,认识临床症状和体征、疾病发展和进展以及治疗策略,绘制与临床检查、病理、生化分析、影像和治疗的关系,整合在一个特定谱学类别中共同指向完整功能的离散元素^[1,8]。Ren 和其同事^[16]在“肿瘤生物信息学主题系列”中介绍了一个名为“分类优化工具”的计算机软件,该软件可应用于例如基因表达特征、蛋白谱等的多种检测,揭示肿瘤的表型、提高诊断的准确性。

肿瘤生物信息学在检测和预计精准医学的效率和有效性中有重要的作用,基于每一个个体基因和蛋白质表达的变异,精准医学提供最安全和有效的

治疗策略。微阵列、蛋白组学、表观基因组学和新一代排序等来源的数据存在语义异质性,通过一个有效的方式整合分子、病理、放射和临床数据,为广泛分布于服务导向、模型来源基础结构的数据库的质疑提供了一个实体的解决方案^[17]。一项最近的研究在人类肝细胞肝癌基因组学分析的前提下进行了基因筛选,通过随后的小鼠模型分析和RNA干扰分析,发现肝癌中一个普遍的基因改变(11q13.3扩增)激活成纤维细胞生长因子19(fibroblast growth factor, FGF19),导致对FGF19抑制剂的选择敏感性^[18]。基于每一个患者肿瘤的分子网络特征,以期可以发展出一种精确的工具,在正确的时间为正确的患者提供正确的治疗。肿瘤生物信息学和系统生物学通过治疗设计,以期改善肿瘤的预防、诊断和治疗情况。分析基因组、生物序列、大规模谱学数据和蛋白质三维结构的经典统计学技术是计算机辅助肿瘤研究中不可或缺的^[19]。

总体而言,肿瘤生物信息学是一种新兴的手段,是系统临床医学中进行临床研究和应用并且改善肿瘤患者预后的一种最关键和最有用的方法。“肿瘤生物信息学主题系列”为科学家整合谱学、生物信息学工具和数据、临床研究、疾病特异性生物标志物、动态网络与精准医学提供了一个独特的平台和难得的机遇,以期共同对抗肿瘤,改善肿瘤患者的生活质量。

【参考文献】

- [1] Wang X, Liotta L. Clinical bioinformatics: a new emerging science[J]. *J Clin Bioinforma*, 2011, 1(1):1.
- [2] Holford ME, McCusker JP, Cheung KH, et al. A semantic web framework to integrate cancer omics data with biological knowledge[J]. *BMC Bioinformatics*, 2012, 13(Suppl 1):S10.
- [3] Ebbert MT, Bastien RR, Boucher KM, et al. Characterization of uncertainty in the classification of multivariate assays: application to PAM50 centroid-based genomic predictors for breast cancer treatment plans[J]. *J Clin Bioinforma*, 2011, 1:37.
- [4] Laczny C, Leidinger P, Haas J, et al. miRTrail - a comprehensive webserver for analyzing gene and miRNA patterns to enhance the understanding of regulatory mechanisms in diseases[J]. *BMC Bioinformatics*, 2012, 13:36.
- [5] Kimori Y. Mathematical morphology-based approach to the enhancement of morphological features in medical images[J]. *J Clin Bioinforma*, 2011, 1:33.
- [6] Haustein V, Schumacher U. A dynamic model for tumour growth and metastasis formation[J]. *J Clin Bioinforma*, 2012, 2(1):11.
- [7] von der Heyde S, Beissbarth T. A new analysis approach of epidermal growth factor receptor pathway activation patterns provides insights into cetuximab resistance mechanisms in head and neck cancer[J]. *BMC Med*, 2012, 10:43.
- [8] Wang X. Role of clinical bioinformatics in the development of network-based Biomarkers[J]. *J Clin Bioinforma*, 2011, 1(1):28.
- [9] Wu D, Zhu D, Xu M, et al. Analysis of transcriptional factors and regulation networks in patients with acute renal allograft rejection[J]. *J Proteome Res*, 2011, 10(1):175-181.
- [10] Chen H, Song Z, Qian M, et al. Selection of disease-specific biomarkers by integrating inflammatory mediators with clinical informatics in AECOPD patients: a preliminary study[J]. *J Cell Mol Med*, 2012, 16(6):1286-1297.
- [11] Cun Y, Fröhlich H. Prognostic gene signatures for patient stratification in breast cancer-accuracy, stability and interpretability of gene selection approaches using prior knowledge on protein-protein interactions [J]. *BMC Bioinformatics*, 2012, 13(1):69.
- [12] Chen H, Wang Y, Bai C, et al. Alterations of plasma inflammatory biomarkers in the healthy and chronic obstructive pulmonary disease patients with or without acute exacerbation[J]. *J Proteomics*, 2012, 75(10):2835-2843.
- [13] Hood L, Friend SH. Predictive, personalized, preventive, participatory (P4) cancer medicine[J]. *Nat Rev Clin Oncol*, 2011, 8(3):184-187.
- [14] Hood L, Heath JR, Phelps ME, et al. Systems biology and new technologies enable predictive and preventative medicine[J]. *Science*, 2004, 306(5696):640-643.
- [15] Tian Q, Price ND, Hood L. Systems cancer medicine: towards realization of predictive, personalized and participatory (P4) medicine[J]. *J Intern Med*, 2012, 271(2):111-121.
- [16] Ren X, Wang Y, Wang J, et al. A unified computational model for revealing and predicting subtle subtypes of cancers[J]. *BMC Bioinformatics*, 2012, 13(1):70.
- [17] González-Beltrán A, Tagger B, Finkelstein A. Federated ontology-based queries over cancer data[J]. *BMC Bioinformatics*, 2012, 13(Suppl 1):S9.
- [18] Sawey ET, Chanrion M, Cai C, et al. Identification of a therapeutic strategy targeting amplified FGF19 in liver cancer by Oncogenomic screening[J]. *Cancer Cell*, 2011, 19(3):347-358.
- [19] Nagl S. Cancer bioinformatics: from therapy design to treatment[M]. Publisher: John Wiley & Sons, 2006:287.