

· 药咖论坛 ·

INSIDERS FORUM



· 评述 ·

生物大数据在精准诊疗药物开发中的应用进展

张鹏程

(上海科技大学生物医学工程学院, 上海 201210)

[中国分类号] R91 ; R318

[文献标志码] A

[文章编号] 1001-5094 (2025) 12-1053-03

DOI: 10.20053/j.issn1001-5094.202512050973

1 精准诊疗药物开发：药物研究的重要方向

随着对疾病分子机制认识的不断深化及临床疗效要求的逐步提升,传统“千人一方”的医疗模式在临床实践中已逐渐显现局限性。不同患者群体乃至个体在遗传背景、分子特征及药物反应性上存在显著差异,此类差异不仅会影响药物治疗效果,还可能导致严重不良反应。因此,基于上述个体差异制定针对性治疗策略,即精准医疗,已成为现代医学发展的重要需求。

精准医疗的实现,依赖于靶点发现、患者分型与针对性药物开发的协同推进^[1]。具体而言,首先通过系统化解析患者分子特征,识别调控疾病发生发展的关键驱动因子;其次利用特异性试剂或探针对患者进行分子分层,明确最佳治疗方案;最后针对可干预的分子靶点,开发特异性小分子药物、抗体药物乃至细胞疗法等治疗手段。

精准医疗已显著提升肿瘤等重大疾病的临床诊治成效。在肿瘤领域,基于基因组分析指导的靶向治疗已成为精准医疗的典型应用;而药物基因组学的发展,使得预测药物代谢差异成为可能,进一步提升了药物治疗的有效性与安全性。与此同时,通过捕获蛋白质表达、翻译后修饰和蛋白-蛋白相互作用

(protein-protein interactions, PPIs)的时空动态特征,蛋白质组学也为生物标志物发现、临床决策制定和个性化医疗实施提供了重要支撑^[2]。

近年来,包括生物样本库数据、电子健康记录在内的医学健康数据持续快速增长,但如何从海量异质性多源数据中精准筛选可干预靶点,并高效开发先进诊断技术和安全有效的治疗药物,仍是当前精准诊疗药物研究领域的前沿难题。

2 生物大数据在靶点发现与诊疗药物开发中的应用进展

在精准诊疗研究中,生物大数据已成为不可或缺的基础支撑条件。生物大数据不仅涵盖多组学信息,还包含与药物研发密切相关的化合物数据库及医学影像数据等,可为靶点发现、探针设计及药物开发全流程提供全方位的数据支撑。

多组学数据包含基因组、转录组、蛋白质组和代谢组数据,分别能够提供遗传变异、基因表达、蛋白功能及代谢通路的动态信息,为疾病驱动因子识别及靶点验证提供多层次证据。通过整合不同研究中的多组学数据并开展统计分析,可挖掘潜在的新药物作用靶点^[3]。例如,Zhang等^[4]近



【本期栏目主编】张鹏程:上海科技大学副教授。2011年毕业于中国科学院上海药物研究所,获药剂学博士学位;2014年从美国约翰霍普金斯大学学成归国,入职上海药物研究所,历任副研究员、研究员;2023年正式加盟上海科技大学。长期聚焦多肽类诊疗药物与递送系统的基础与转化研究。科研项目方面,先后获上海市浦江人才计划(2015)、中国科学院青年拔尖人才(2023)等人才项目资助;入选中国科学院青年创新促进会会员(2017),并获评该促进会优秀会员(2021)。

张鹏程课题组深耕医学诊疗用多肽领域,围绕多肽靶分子、多肽药物及多肽载体三大方向,系统开展多肽高通量设计、合成与筛选技术研发,攻关多肽键合修饰与超分子组装关键技术,阐明多肽-生物体相互作用的分子机制,最终实现技术成果在肿瘤等重大疾病诊断与治疗中的转化应用,取得系列创新性成果。课题组已在*Nat Nanotechnol*、*Nat Commun*、*Sci Adv*、*Adv Mater*、*J Control Release*、*Acta Pharm Sin B*等权威SCI期刊发表学术论文40余篇。

期在 *Science* 报道了一种基于人工智能 (artificial intelligence, AI) 的 PPIs 预测方法, 该方法通过深度学习结构预测模型整合超大规模基因组与蛋白质组数据, 对约 2 亿个人类蛋白对进行系统筛查, 最终预测出 17 849 个高置信度 PPIs, 并构建了相应的三维复合物结构, 其中有 3 631 个 PPIs 此前未被实验验证发现。

生物大数据的应用还显著提升了疾病分子诊断的效率与深度。Lunke 等^[5]开展了整合多组学数据的临床研究, 通过将超快速全基因组测序与转录组分析、蛋白质组分析及功能验证相结合, 平均可在 3 天内完成危重儿童罕见遗传病的基因诊断报告, 整体诊断率提升至 54%。进一步整合计算机断层扫描、磁共振成像、正电子发射断层显像及超声等医学影像数据, 能够在原有分子标志物量化信息的基础上, 进一步提供其空间分布、组织结构等关键信息, 从而提升诊断准确性并助力疗效预测。基于上述策略, Zhang 等^[6]近期通过整合乳腺癌靶点与超声影像、临床数据及其他多模态信息, 构建 AI 模型, 实现了对患者人表皮生长因子受体 2 状态分型及其新辅助治疗反应的精准预测。Chen 等^[7]通过整理 32 个器官的 220 万对配对数据集, 构建视觉-组学基础模型 OmiCLIP, 实现了组织病理图像与空间转录组学数据的跨模态桥接与整合分析, 为疾病分子分型与针对性治疗方案制定奠定了基础。

抗体药物及其衍生物研发近年来呈快速增长态势, 其中双特异性抗体、多特异性抗体凭借其多价结合与桥联特性, 在肿瘤治疗中展现出独特优势, 已成为当前抗体药物研发的热点方向^[8]。传统抗体药物验证阶段需经过抗原制备、动物免疫、抗体筛选与制备等多个流程, 存在耗时久、成本高的问题, 而目标抗原组合选择与特异性识别域开发则是该领域的一大挑战。借鉴生物大数据驱动小分子药物研发的技术路径, 通过整合蛋白质组学数据、结构生物学数据和化合物-蛋白结合信息, 研究者可实现结合亲和力的预测, 进而指导抗体分子构建。Bennett 等^[9]近期在 *Nature* 报道了一种新型抗体设计技术, 该技术将微调后的 RFdiffusion 网络蛋白虚拟设计与酵母展示筛选相结合, 基于多源生物大数据中抗体-抗原配对样本进行训练, 实现了对抗体结合模式的原子级精度设计及对特定表位的精准靶向。同时, 生物大数据还可辅助优化表位组合选择, 评估抗体的多靶点效应及潜在副作用, 为双特异性或多特异性抗体的药效优化提供理论依据。

近年来, AI 技术在药物研发的靶点验证、探针设计及药物优化等关键环节的深度系统应用,

为生物大数据进一步整合及药物研发的自动化推进提供了可能。Xu 等^[10]近期在 *Nature Medicine* 报道了首个由生成式 AI 同时完成靶标识别和化合物设计并进入临床试验的研究成果, 成功识别出 TRAF2 和 NCK 相互作用激酶作为特发性肺纤维化的关键靶点, 并基于 AI 模型设计出小分子抑制剂 rentosertib, 该药物在 IIa 期临床试验中展现出良好的安全性和改善肺功能的潜在疗效。

3 本期专题论文点评

本期专题精选 3 篇聚焦生物大数据驱动精准诊疗药物研究的论文, 系统呈现该领域的最新进展与创新思路。其中, 上海科技大学白芳团队系统综述了生物大数据的类型与资源及其在靶标发现与验证、先导化合物发现与优化两大环节的应用进展, 并提出生物大数据可溯源性、预测逻辑的可解释性及预测与实验闭环的可进化性, 将是未来精准药物开发的潜在突破方向。中国科学院宁波材料技术与工程研究所吴爱国团队提出肿瘤边界生物学理念, 重点阐述了基于乳腺癌病灶边界生物组学信息开发的纳米级医学影像造影剂的设计原理、技术思路与应用进展; 同时指出, 边界生物组学信息的整合、安全特异性探针的规模化制备及配套成像装备的临床普及, 将是该领域未来的研究重点。华南理工大学王均和沈松团队总结了生物互作数据 (包括细胞互作、受体互作、多价互作) 启发的双/多特异性抗体设计及制备策略, 深入梳理了基于生物大数据的 AI 驱动设计技术及纳米技术在双/多特异性抗体研发 (涵盖靶点预测、靶点组合筛选、成药性预测) 中的重要进展, 并提出提升生物数据质量及 AI 算法的准确性与可解释性, 有望推动双/多特异性抗体药物研发领域的突破性进展。

4 结语与展望

生物大数据在诊疗药物研发中的应用已取得显著进展。通过整合多组学数据、化合物数据库及医学影像数据等, 研究者可高效完成靶点识别、分子探针设计, 并实现小分子药物与抗体药物的优化开发。然而, 该领域仍面临诸多挑战: 其一, 异质性多源生物大数据 (涵盖数据类型、标准及质量的差异) 的有效整合与合理应用, 仍是当前的核心难点; 其二, 自动化、高通量的靶点验证、分子设计与合成及体内活性验证技术尚未完全成熟, 限制了研发效率的提升; 其三, 如何将生物大数据有效应用于药物临床试验设计 (包括患者分层、临床终点设定等), 最终实现临床获益最大化, 仍需进一步探索。未来, 随着 AI 加速的数

据产出和分析、AI 驱动的高通量实验室技术及 AI 辅助的临床试验设计的协同发展，生物大数据在药学研

究中的潜力将得到进一步释放，有望实现靶点、探针和药物的一体化开发，为精准药物研发提供坚实支撑。

【参考文献】

- [1] Wong E, Bertin N, Hebrard M, *et al.* The Singapore national precision medicine strategy[J]. *Nat Genet*, 2023, 55(2): 178–186.
- [2] Jiang Y, Wang J, Sun A, *et al.* The coming era of proteomics-driven precision medicine[J]. *Natl Sci Rev*, 2025, 12(8): nwaf278.
- [3] Santos A, Colaço A R, Nielsen A B, *et al.* A knowledge graph to interpret clinical proteomics data[J]. *Nat Biotechnol*, 2022, 40(5): 692–702.
- [4] Zhang J, Humphreys I R, Pei J, *et al.* Predicting protein-protein interactions in the human proteome[J]. *Science*, 2025, 390(6771): eadt1630.
- [5] Lunke S, Bouffler S E, Patel C V, *et al.* Integrated multi-omics for rapid rare disease diagnosis on a national scale[J]. *Nat Med*, 2023, 29(7): 1681–1691.
- [6] Zhang J, Li Y, Li Z, *et al.* Deep-learning-based HER2 status assessment from multimodal breast cancer data predicts neoadjuvant therapy response[J/OL]. *Nat Biomed Eng*, 2025[2025-12-04]. <https://doi.org/10.1038/s41551-025-01495-5>.
- [7] Chen W, Zhang P, Tran T N, *et al.* A visual-omics foundation model to bridge histopathology with spatial transcriptomics[J]. *Nat Methods*, 2025, 22(7): 1568–1582.
- [8] Klein C, Brinkmann U, Reichert J M, *et al.* The present and future of bispecific antibodies for cancer therapy[J]. *Nat Rev Drug Discov*, 2024, 23(4): 301–319.
- [9] Bennett N R, Watson J L, Ragotte R J, *et al.* Atomically accurate *de novo* design of antibodies with RFdiffusion[J/OL]. *Nature*, 2025[2025-12-04]. <https://pubmed.ncbi.nlm.nih.gov/41193805/>. DOI: 10.1038/s41586-025-09721-5.
- [10] Xu Z, Ren F, Wang P, *et al.* A generative AI-discovered TNIK inhibitor for idiopathic pulmonary fibrosis: a randomized phase 2a trial[J]. *Nat Med*, 2025, 31(8): 2602–2610.

（责任编辑：邢爱敏）