

基于疾病演变分子轨迹发现中药作用靶标及先导化合物思路探讨

张瑜¹, 龙江兰¹, 王爱婷¹, 吕昊², 邓科君^{2*}, 林昊^{2*}, 鄢丹^{1*}

(1. 首都医科大学附属北京友谊医院, 北京市临床药学研究所, 北京 100050;

2. 电子科技大学生命科学与技术学院, 四川成都 610054)

摘要: 探索中药作用靶标(群)是推动中药创新发展的重要命题,但囿于“是否与疗效相关、与疾病相关”而备受关注。项目组前期通过临床研究发现,糖尿病发生发展中的代谢组特征分子,与玉泉丸疗效应答显著相关。以此为线索,拟拓展糖尿病发生发展的多组学特征信息,发现玉泉丸降糖关键作用靶标(群)及其先导化合物。包括:①从临床切入,运用融入生成式人工智能的组学技术,挖掘蛋白质组、微生物组特征信息,连同代谢组特征分子整合形成驱动因素,刻画糖尿病演变分子轨迹及其经玉泉丸治疗后的转归;②以演变分子轨迹为纽带与指针,运用拟人源化模型和化学蛋白质组等分子生物学技术,发现玉泉丸降糖关键作用靶标(群),配以前瞻性临床样本验证;③利用图神经网络技术建立关键作用靶标(群)整体评判方法,探寻临床病证明确、作用机制清晰的药源性/内源性先导化合物,为关联临床疗效的中药复杂活性成分作用靶标(群)与创新药物的发现提供新范式、新技术。

关键词: 关联临床; 糖尿病演变分子轨迹; 作用靶标; 先导化合物; 玉泉丸

中图分类号: R917 文献标识码: A 文章编号: 0513-4870(2024)08-2199-06

Discovery of the targets and lead compounds of traditional Chinese medicine based on the molecular trajectory of diabetes evolution

ZHANG Yu¹, LONG Jiang-lan¹, WANG Ai-ting¹, LÜ Hao², DENG Ke-jun^{2*}, LIN Hao^{2*}, YAN Dan^{1*}

(1. Beijing Institute of Clinical Pharmacy, Beijing Friendship Hospital, Capital Medical University, Beijing 100050, China; 2. School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu 610054, China)

Abstract: Exploring the action targets (groups) of traditional Chinese medicine (TCM) is an important proposition to promote the innovation and development of TCM, but it has attracted a lot of attention as to whether it is related to the efficacy or the disease. Our team found that the metabolomic signature molecules in the development of diabetes mellitus (DM) were significantly associated with the clinical efficacy of Yuquan Pill through a large clinical sample study. Taking this as a clue, our team intends to expand the information on the omics features of DM development, and discover the key targets (groups) and their lead compounds for the hypoglycemic effect of Yuquan Pill. The project includes: ① Based on the retrospective clinical trials, using omics technology integrated with generative artificial intelligence, mining the characteristic information of proteome and microbiome, forming driving factors together with metabolome characteristic molecules, and characterizing the molecular trajectories of diabetes evolution and their interference by Yuquan Pill; ② Taking the evolving molecular

收稿日期: 2024-06-03; 修回日期: 2024-07-03.

基金项目: 国家自然科学基金项目 (82130112, 82204699); 青年北京学者项目 (2022-051); 国家中医药管理局高水平中医药重点学科建设项目 (zyyzdxk-2023003).

*通讯作者 Tel: 86-28-83208238, E-mail: dengkj@uestc.edu.cn;

Tel: 86-28-83208238, E-mail: hlin@uestc.edu.cn;

Tel: 86-10-63139318, E-mail: danyan@ccmu.edu.cn

DOI: 10.16438/j.0513-4870.2024-0526

trajectories as a link and pointer, using anthropomorphic modeling and molecular biology techniques such as chemical proteomics to discover the key targets (groups) of Yuquan Pill's hypoglycemic effect, with the prospective clinical samples for validation; ③ Evaluate the overall response of key targets (groups) using graph neural network technology, and search for drug-derived/endogenous lead compounds with proven clinical pathologies and clear mechanisms of action, so as to provide a new paradigm and technology for the discovery of complex active ingredient targets (groups) of TCM that are related to their clinical efficacy, as well as for the discovery of innovative medicines.

Key words: related to clinical efficacy; molecular trajectory of diabetes evolution; target; lead compound; Yuquan Pill

中药是在中医药临床实践中孕育和发展的,探索中药复杂活性成分作用靶标(群),已成为推动中药创新发展的重要命题^[1,2]。然而,发现的作用靶标(群)在阐释中药复杂活性成分生物机制的同时,“是否与疗效相关、与疾病相关”,已成为当前中药复杂活性成分作用靶标(群)研究领域,不可回避的科学问题^[2,3]。

面对中药复杂活性成分作用靶标(群),“是否与疗效相关、与疾病相关”的难题,本文秉持“源于临床—证于试验—归于临床”研究策略^[4],本着中药源自临床医疗实践的初衷,以治疗2型糖尿病临床疗效确切的玉泉丸为例:①从前期开展的回顾性匹配队列切入,挖掘疾病发生发展的特征分子(驱动因素),刻画疾病演变分子轨迹及其经中药治疗后的转归;②以此为纽带与指针,联通临床与基础,探索中药对疾病演变分子轨迹的转归机制,发现中药复杂活性成分的关键作用靶标(群),配以前瞻性临床样本验证;③立足关键作用靶标(群)整体应答,探寻潜在临床价值突出、临床适宜病证明确、主要作用机制清晰的先导化合物,以推动关联临床疗效的中药复杂活性成分作用靶标(群)的发现及其创新药物的研发。这对中药复杂活性成分作用靶标(群)的有效发现至关重要,也是释除当前中药复杂活性成分的作用靶标(群)难关疗效、难关疾病而备受争议的有效举措。

1 挖掘糖尿病发生发展的特征分子,刻画糖尿病演变分子轨迹及其经玉泉丸治疗后的转归

疾病发生发展是一项复杂生物学过程,涉及基因、蛋白质、微生物、代谢物等交互参与,临床诊断结局指标尚难以反映其复杂信息。立足“分子整体状态医学”,遵从中医药循证医学思想,利用组学技术挖掘中医药优势病种发生发展的特征分子,以此为驱动因素刻画疾病演变分子轨迹;为进一步通过评估中药对疾病演变分子轨迹的应答与转归,发现关联临床疗效的中药活性成分作用靶标(群)提供了重要突破口^[5]。

糖尿病属于中医“消渴”范畴,已被列入中医药优势病种。玉泉丸为治疗消渴的经典名方,出自清代名

医叶天士《种富堂公选良方》,由葛根、天花粉、地黄、麦冬、五味子、甘草组方,具有养阴生津、止渴除烦、益气和中功效^[6],用于气阴不足、口渴多饮、消食善饥,临床可用于治疗上述症状的糖尿病患者。

挖掘糖尿病发生发展特征分子,刻画糖尿病演变分子轨迹及其经玉泉丸治疗后的转归:拟获取回顾性匹配队列研究中的糖尿病前期和2型糖尿病,以及经玉泉丸治疗后的血液、粪便等样本;融入生成式人工智能的多组学技术,抽提代谢组、蛋白质组、微生物组反映疾病发生发展不同阶段和玉泉丸治疗前后/有效/无效的生物特征信息,作为驱动因素,刻画糖尿病演变分子轨迹及其经玉泉丸治疗后的转归。

1.1 建立研究用临床生物样本和知识图谱 从临床入手,在项目组已完成1 507 563例糖尿病人群横断面、387 076例长达10年随访研究的基础上^[7],开展前瞻性队列研究,招募全国5个临床中心1 705例大样本(糖尿病前期组509例、2型糖尿病组555例、健康组495例、干扰组146例),以及糖尿病前期、2型糖尿病人群(中医消渴气阴两虚证)拟经玉泉丸治疗后的血液、粪便样本和人口学、生命体征、实验室生化指标、中医证候、西医疗效等信息,连同健康组来源的血液、粪便、人口学等作为参照组,可形成研究用临床生物样本和知识图谱,包括糖尿病发生发展不同阶段横断面和玉泉丸治疗前后/有效/无效纵向随访人群。

1.2 融入生成式人工智能的多组学技术,挖掘糖尿病发生发展特征分子 按“发现—测试—验证”三阶段设计实验,由单中心向多中心推演验证,利用代谢组、蛋白质组、微生物组技术,抽提关联玉泉丸降糖疗效的生物特征信息,借助生成式人工智能探索多组学大数据内在逻辑关联(驱动因素),发现糖尿病发生发展特征分子。

研究团队前期已开展了全国5个临床中心1 705例大样本临床代谢组分析(2型糖尿病组555例、糖尿病前期组509例、健康组495例、干扰组146例,中国临床试验注册号ChiCTR1800014301)^[8,9]。发现内源分子乙

酰基左旋肉碱、L-异亮氨酸、L-精氨酸、左旋肉碱、L-谷氨酸、L-亮氨酸、L-赖氨酸、L-蛋氨酸、溶血磷脂酰胆碱(P-16:0)和L-苯丙氨酸,与糖尿病前期、2型糖尿病发生发展显著相关,为探寻基于代谢组的玉泉丸临床疗效应答指标提供了重要线索^[10,11]。

1.3 基于特征分子(驱动因素)时序变化,刻画糖尿病演变分子轨迹及其经玉泉丸治疗后的转归 以抽提出的多组学生物特征信息为驱动因素,采用GINI指数和互信息分析,通过特征去噪机制,降低多模态数据分析中可能引入的数据噪声,精确识别与疾病发生发展相关的特征分子。进一步地,以特征分子(驱动因素)为基点,针对糖尿病发生发展不同阶段和治疗前后/有效/无效,开展横向关联、纵向时序标准化处理特征分子的时空分布,运用时间序列分析方法(ARIMA模型),评估治疗前后特征分子的变化趋势;运用变化点检测算法(Bai-Perron断点检测),评估治疗引发的特征分子变化时间,从而刻画糖尿病演变分子轨迹及其经玉泉丸治疗后的转归,可用于糖尿病风险预测、演变评估、疗效评判。

2 发现关联临床疗效的玉泉丸降糖关键作用靶标(群)

将源于临床发现的疾病演变分子轨迹(驱动因素)及转归,通过迁移学习、因果学习映射到动物、细胞水平研究中,结合拟人源化模型的分子生物学(人源化动物、细胞、3D类器官,以及空间蛋白质组、空间转录组等)^[12,13],明确糖尿病演变驱动因素的组织来源、产生过程,以及中药复杂活性成分作用靶标(群),实现由临床表型向深达生物机制的阐释,实现“临床与基础”联通。

2.1 建立关联临床疗效的生物机制研究评价指针 发挥系统药理学具有整合时空尺度数据的模拟分析能力,发现药物成分与机体之间潜在的多层多维关系(药物分子、靶标、细胞、组织、器官等)优势,利用GSEA数据库MSigDB对潜在作用靶标进行注释富集,剖析玉泉丸可能的活性成分与糖尿病相关靶标(蛋白激酶、核受体蛋白、G蛋白偶联受体等),涉及Cytoscape3.9.0、MetaboAnalyst5.0、KEGG pathway等工具;结合网络拓扑结构和贡献指数综合分析,发现与玉泉丸治疗糖尿病相关的信号通路和靶标的器官分布,构建“成分—靶标、靶标—通路和靶标—器官”网络。

2.2 基于动物模型、细胞模型,评估药效、药源性成分代谢转化特征 基于db/db自发性糖尿病小鼠、高脂饮食诱导的糖尿病前期小鼠模型、高糖高脂饲料联合低剂量链脲佐菌素诱导的2型糖尿病大鼠模型和人源化糖尿病模型小鼠评估玉泉丸药效;体内分析玉泉丸成

分代谢转化特征:探究药源性成分代谢转化特征,明确其主要代谢产物;基于胰岛细胞凋亡模型、大鼠原代肝细胞模型、脂肪细胞胰岛素抵抗模型、HepaRG细胞3D类器官模型评估玉泉丸的降糖活性:从细胞活性、细胞损伤、胰岛素分泌功能三方面考察玉泉丸肠道菌混悬液、含药血清对不同细胞模型的药效和机制作用。

2.3 迁移临床评价指针发现玉泉丸活性成分降糖的关键作用靶标(群) 开展上述实验样本的代谢组、蛋白质组、微生物组及其关键产物、酶等特征信息(驱动因素)靶向检测,通过迁移学习、因果学习技术,建立的“关联临床疗效的基础端生物机制研究评价指针”进行衡量,辅以必要的空间组学技术挖掘作用靶标(群);基于前期系统药理学初探研究和蛋白质组小样本预试验的发现,重点围绕降糖活性经典信号通路胰岛素信号通路(insulin-IRS-PI3K-Akt)、AMPK信号通路(AMPK-PEPCK等)、mTOR信号通路(mTOR-S6K1-IRS-1),以及免疫球蛋白lambda常数3(immunoglobulin lambda constant 3, IGLC3)、免疫球蛋白重常数 μ (immunoglobulin heavy constant mu, IGHM)、免疫球蛋白 κ 常数(immunoglobulin kappa constant, IGKC)的丢失与N-糖基化修饰等开展研究^[14],主要组织涉及肝脏、脂肪等,并结合目标蛋白/基因/酶过表达/稳定干涉和Co-IP、pull-down等分子生物学技术,发现玉泉丸降糖活性关键作用靶标(群)及其主要信号通路等。

基于空间代谢组、空间蛋白质组及空间转录组分析:揭示糖尿病代谢组特征分子的组织分布、代谢变化轨迹、关联靶标蛋白,及其经玉泉丸治疗后的代谢调控机制等提供重要研究线索;揭示糖尿病特征蛋白质的组织分布、变化轨迹、关键靶标蛋白,及其经玉泉丸治疗后的关键靶标蛋白变化规律,结合过表达/稳定干涉等分子生物学研究,推动玉泉丸降糖活性关键作用靶标(群)的发现及其生物机制阐释;进一步佐证发现玉泉丸降糖活性关键作用靶标(群)及其生物机制。

2.4 回归临床前瞻性样本,验证玉泉丸降糖活性的关键作用靶标(群) 取糖尿病前期、2型糖尿病组以及经玉泉丸治疗有效受试者;另按相同入组/排除标准新招募糖尿病前期、2型糖尿病患者(分为两组,一组为玉泉丸治疗组,另一组为参照组),分别收集其血液、粪便样本,采用靶向代谢分析、Western blot、ELISA技术,对上述研究中发现的关键作用靶标(群)进行检测验证。

以系统药理学初探为可能的线索、实验研究数据为发现、临床验证数据为确认,将上述降糖相关信号通路中关键蛋白、小分子与在临床样本中验证的结果进行对比,相互印证确认关联临床疗效的玉泉丸降糖活性关键作用靶标(群)。

3 探寻临床病证明确、作用机制清晰的先导化合物

基于关键作用靶标(群)整体应答,探寻药源性/内源性先导化合物:以药效应答或转归为依据,将关键作用靶标(群)作为Panel进行整体应答评价,利用拟人源化模型以及Transformer深度学习技术等筛选降糖活性潜在化合物,辅以人工智能优化设计,开展类药性、有效性、安全性和成药性研究,以探寻潜在临床价值突出、临床适宜病证明确、主要作用机制清晰的先导化合物。

3.1 建立基于关键作用靶标(群)整体应答评判方法,辅以自主研发的降糖药物智能设计平台筛选先导化合物 基于代谢组特征分子参与对玉泉丸降糖活性应答的事实,拓展到上述多组学与分子生物学研究发现的关键作用靶标(群),据此发现源于玉泉丸治疗发挥过程中的药源性/内源性先导化合物。将关键作用靶标(群)作为Panel,以GINI系数和互信息技术明确关键作用靶标(群)有效信息的逻辑关联,利用Transformer图神经网络技术建立图像化的关键作用靶标(群)整体应答评判方法,用于先导化合物降糖活性的筛选。

项目组前期依据国际降糖药物研发领域的“结构-活性-靶标”知识图谱,运用机器学习技术(分子指纹编码+增量特征选择),建立的降糖药物智能设计平台,可用于本研究中发现的具有潜在成为降糖活性先导化合物的目标进行预测评价(<https://github.com/lllxw/iPADD>)。具体地,基于空间多组学数据信息,利用自注意力机制捕捉序列长程依赖及潜在的功能域信息,生成具有潜在活性的候选化合物;利用深度强化学习(Molecular RNN和GraphGA)进行化合物结构优化,通过迭代学习改善化合物药理学属性和合成可行性;采用标准评价指标ROC-AUC对化合物的合成可行性和生物相容性评分评判。

3.2 基于关键作用靶标(群)整体应答为指引,探寻具有降糖活性潜力的先导化合物 以玉泉丸降糖活性及生物机制研究中,差异性的药源性原型成分及代谢产物、内源性成分(代谢小分子、蛋白质、微生物及其关键酶等)为候选化合物,通过降糖药物智能设计平台进行预测;利用表面等离子共振(surface plasmon resonance, SPR)技术靶向垂钓筛选潜在先导化合物,将关键作用靶标(群)中显著应答蛋白(Western blot、ELISA检测)固定在芯片上,制作系列靶标芯片,分析潜在先导化合物与靶标芯片的特异性、亲和力、结合模型、再生条件、动力学等;利用降糖细胞/动物模型,以关键作用靶标(群)整体应答为指引(结合降糖结局指标),发现降糖活性先导化合物。

以前述研究中发现的玉泉丸降糖生物机制,开展

相关信号通路上的代谢小分子、蛋白质、微生物及其关键酶等靶向检测,结合过表达/稳定干涉等常规分子生物学技术,阐释先导化合物降糖生物机制。

4 讨论

当前中药临床疗效主要由中医非特异症状群表示,存在主观性和精准性不足等局限;疾病演变分子轨迹具有反映疾病分子整体状态及药物干预的特点,与主要疗效指征、证候指征、生物活性、作用机制关联,从而为通过疾病演变分子轨迹反映中药临床疗效提供了生物学可能。融入多组学特征信息的疾病演变分子轨迹是包容中医证候、西医疗效、系统生物学等指征对疾病整体状态的表达,在生物机制层面,体现整体活性与生物网络的结合、作用靶标与特定通路的结合;在中药疗效层面,体现药味组方与药效物质的结合、原型成分与代谢成分的结合。

面对源于临床的多组学生物信息与病证结局指标等多维度大数据,如何发掘可关联疾病发生发展与转归的特征数据,构建生物网络清晰、逻辑推理合理、量化精准辨识的“分子整体状态医学”评价指征,以推动中药复杂活性成分作用靶标(群)发现与作用机制阐释,是当前该研究领域的技术难题。人工智能AI具有智能学习、逻辑推理、信息感知、辨识预测、可解释性等能力,适合处理多重复杂的生物医学大数据信息,智能解析复杂生物学事件,注解核酸、蛋白质、代谢物、微生物等与临床疗效之间的关联关系^[9,15]。特别是近年来,作为人工智能迭代的核心驱动力——生成式人工智能Generative AI不仅可以分析或预测数据,更是可以通过循环神经网络RNN、生成对抗网络GAN、变分自编码器VAE、转换器Transformer等概率模型或神经网络模型,可解释性地实现文本数据、图像数据、视频数据之间的逻辑关联与逻辑生成,有利于厘清疾病多模态特征分子之间的网络关系,推动疾病跨模态特征分子的生成式发现^[16,17]。因此,生成式人工智能为从繁杂的组学大数据中,挖掘疾病发生发展特征分子,刻画疾病演变分子轨迹提供了技术可能^[18]。

探寻临床病证明确、作用机制清晰的先导化合物,是创新药物研发的重要源头。本文为从中药复方中发现其作用靶标及先导化合物提供一种研究思路,并以玉泉丸和消渴为示例进行介绍(图1)。重在发现先导化合物,再研究其有效性、安全性、类药性、成药性等,进而为研发临床定位明确、作用机制清晰的创新药提供线索;在当前探寻先导化合物的众多途径中,源于临床重要表型不失为一种多快好省的可靠途径。后续仍需要关注并回答两方面问题:①体现疾病发生发展的特征分子网络如何映射到基础研究中,以实现“临床与

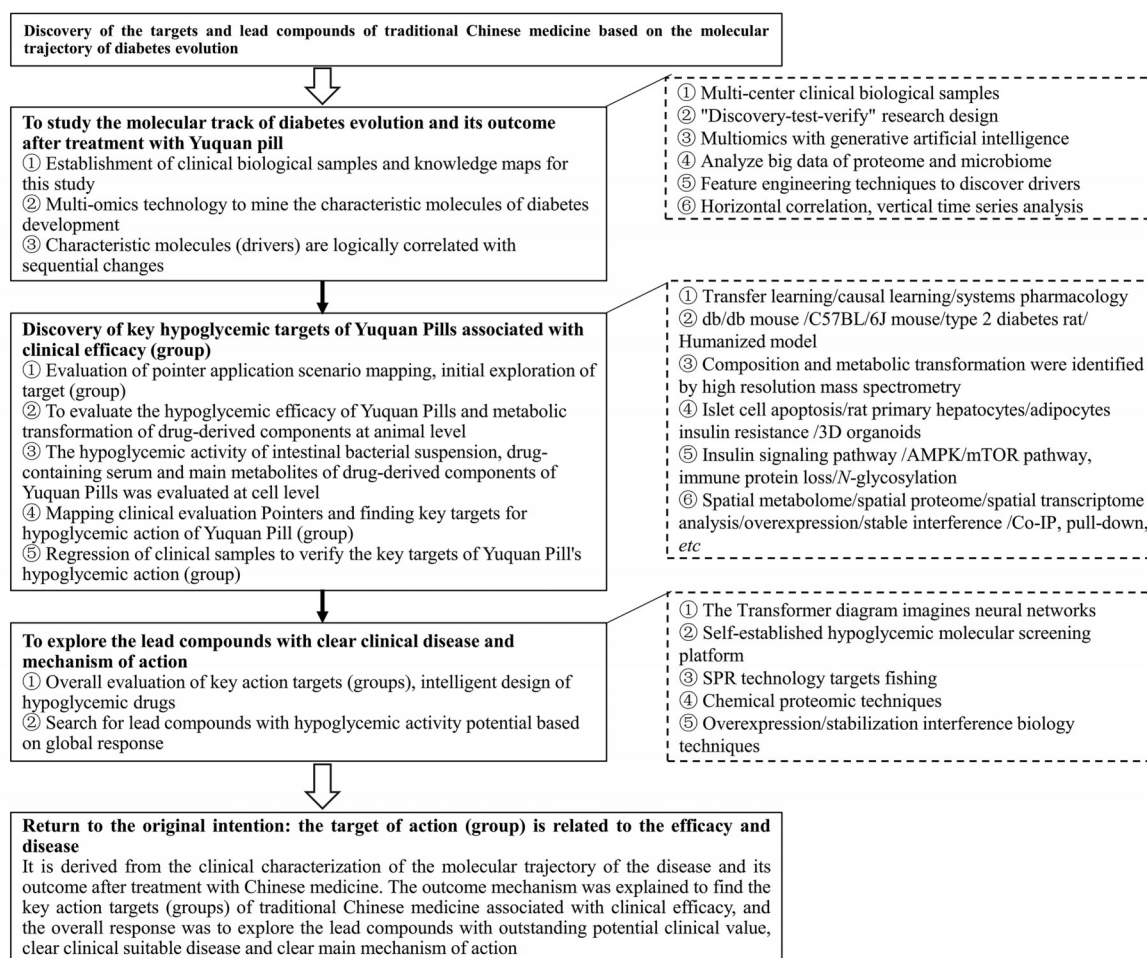


Figure 1 Discovery of the targets and lead compounds of Yuquan Pill based on the molecular trajectory of diabetes evolution

基础”联通; ② 关联临床疗效的作用靶标 (群) 整体应答如何评估, 以体现“分子网络”整体效应。

作者贡献: 张瑜负责文献调研和文章撰写; 龙江兰、王爱婷、吕昊、邓科君和林昊负责文章修改; 鄢丹负责文章框架设计和修改。

利益冲突: 所有作者和所涉及的内容不存在潜在的利益冲突。

References

- [1] General Office of the State Council. 14th five-year plan for the development of traditional Chinese medicine [EB/OL]. 2022-03-29. https://www.gov.cn/xinwen/2022-03/29/content_5682270.htm.
- [2] Cheng HB, Zhang L, Fu Y, et al. 2023 major scientific problems, engineering and technical problems and industrial technical problems of traditional Chinese medicine [J]. J Tradit Chin Med (中医杂志), 2023, 64: 1405-1421.
- [3] Xu ZJ, Fang LZ, Pan DZ. Regulate prescription of Chinese medicines [J]. Nature, 2018, 553: 405.
- [4] Yan D, Liu J, Wang AT, et al. Based on the evidence of evidence-based medicine, the mechanism of action of the dominant diseases of traditional Chinese medicine is discussed [J]. China J Chin Mater Med (中国中药杂志), 2018, 43: 2633-2638.
- [5] Xu M, Deng JW, Xu KK, et al. In-depth serum proteomics reveals biomarkers of psoriasis severity and response to traditional Chinese medicine [J]. Theranostics, 2019, 9: 2475-2488.
- [6] Lei Y, Guo XH, Li Y, et al. Research progress of chemical constituents and pharmacological actions of Yuquan pills and predictive analysis of quality markers [J]. Chin Tradit Herb Drugs (中草药), 2022, 53: 2929-2936.
- [7] Yang H, Luo YM, Ren XL, et al. Risk prediction of diabetes: big data mining with fusion of multifarious physical examination indicators [J]. Inf Fusion, 2021, 75: 140-149.
- [8] Yan D, Long JL, Wang AT. Application of *Staphylococcus slowus* in the preparation of compositions: CN, ZL202211146734.1 [P]. 2023-03-24.
- [9] Li X, Yu HC. A new paradigm for AI-driven life science research [J]. Bull Chin Acad Sci (中国科学院院刊), 2024, 39: 50-58.
- [10] Long JJ, Yang H, Yang ZR, et al. Integrated biomarker profiling

- of the metabolome associated with impaired fasting glucose and type 2 diabetes mellitus in large-scale Chinese patients [J]. *Clin Transl Med*, 2021, 11: e432.
- [11] Yan D, Long JL, Yang ZR. Integrated biomarker profiling for assessing the risk of impaired fasting blood glucose and type 2 diabetes mellitus: CN, ZL202110144115.8 [P]. 2021-06-08.
- [12] Wu L, Ai YJ, Xie RX, et al. Organoids/organs-on-a-chip: new frontiers of intestinal pathophysiological models [J]. *Lab Chip*, 2023, 23: 1192-1212.
- [13] Cheng JJ, Liao J, Shao X, et al. Multiplexing methods for simultaneous large-scale transcriptomic profiling of samples at single-cell resolution [J]. *Adv Sci (Weinh)*, 2021, 8: 2101229.
- [14] Zhao Y, Wang M, Meng B, et al. Identification of dysregulated complement activation pathways driven by *N*-glycosylation alterations Q2 in T2D patients [J]. *Front Chem*, 2021, 9: 677621.
- [15] Cui HT, Wang C, Maan H, et al. scGPT: toward building a foundation model for single-cell multi-omics using generative AI [J]. *Nat Methods*, 2024. DOI: 10.1038/s41592-024-02201-0.
- [16] Ding L. Generative Artificial Intelligence – the Logic and Application of AIGC (生成式人工智能: AIGC 的逻辑与应用) [M]. Beijing: CITIC Publishing Group. 2023.
- [17] Zhou P, Wen L, Lin J, et al. Integrated unsupervised-supervised modeling and prediction of protein-peptide affinities at structural level [J]. *Brief Bioinform*, 2022, 23: bbac097.
- [18] Li LJ, Ji PY, Zheng AL, et al. Research progress of artificial intelligence combined with physiological pharmacokinetic model [J]. *Acta Pharm Sin (药学报)*, 2024. DOI: 10.16438/j.0513-4870.2024-0195.