DOI: 10.3969/j.issn.2097-7948.2026.01.010

· 综述 ·

多组学与人工智能整合在血液肿瘤诊疗中的研究进展

孙瑞霞[1,2]，赵邢力[1,2*]
[1]天津市人民医院，南开大学第一附属医院，肿瘤诊疗中心血液科，天津 300121
[2]天津市人民医院，南开大学第一附属医院，中西医结合肿瘤研究所，天津 300121

摘要： 血液肿瘤的生物学行为及临床预后具有较为复杂的异质性，传统基于形态学和有限分子标志物的分类体系难以满足精准诊疗的需求。随着高通量测序、单细胞、空间组学等技术的成熟，涵盖基因组、表观基因组、转录组、蛋白质组与代谢组的多维数据得以系统解析，为揭示克隆演化轨迹、耐药形成路径及免疫生态重塑提供了全新视角。与此同时，机器学习与深度学习等人工智能方法在处理高维、多模态生物医学数据、自动特征提取和构建预测模型方面展现出显著优势，逐渐从辅助分析工具演变为推动诊断、风险分层和疗效预测的核心方法。本文综述了多组学数据特征及人工智能方法原理，系统总结了两者在血液肿瘤分子分型、复发风险评估以及治疗反应预测中的整合应用进展，并分析了当前存在的局限性和面临的关键挑战，旨在为完善血液肿瘤风险分层体系、推动血液疾病精准诊疗提供新的方法学参考与研究方向。

关键词： 血液肿瘤；多组学；人工智能；精准诊疗；预后预测
中图分类号： R733.7

Research advances in integrating multi-omics and artificial intelligence for the diagnosis and treatment of hematological malignancies

SUN Ruixia[1,2], ZHAO Xingli[1,2*]
[1]Department of Hematology, Oncology Center, Tianjin Union Medical Center, The First Affiliated Hospital of NanKai University, NanKai University, Tianjin 300121, China
[2]Tianjin Cancer Institute of Integrative Traditional Chinese and Western Medicine, Tianjin Union Medical Center, The First Affiliated Hospital of Nankai University, NanKai University, Tianjin 300121, China

Abstract: Hematological malignancies are characterized by pronounced molecular heterogeneity and diverse clinical outcomes, rendering traditional classification systems based on morphology and a limited set of molecular markers insufficient for the requirements of precision medicine. With the maturation of high-throughput sequencing, single-cell and spatial omics technologies, multidimensional data encompassing the genome, epigenome, transcriptome, proteome, and metabolome can now be systematically interrogated, providing new insights into elucidating clonal evolutionary trajectories, mechanisms of drug resistance, and the remodeling of the immune microenvironment. In parallel, artificial intelligence (AI) methods,

基金项目： 天津市医学重点学科建设资助（TJYXZDXK-3-003A-5）；天津市中西医结合重点专科建设项目资助（ZDZKKF02）
* 通信作者
利益冲突： 作者均声明无利益冲突
伦理批件： 不需要
知情同意： 不需要
引用本文： 孙瑞霞，赵邢力. 多组学与人工智能整合在血液肿瘤诊疗中的研究进展[J]. 再生医学研究，2026，1(3)：192-203. doi：10.3969/j.issn.2097-7948.2026.01.010.

Fund program: Tianjin Key Medical Discipline Construction Project (TJYXZDXK-3-003A-5); Tianjin Key Specialty Construction Project for Combining Traditional Chinese and Western Medicine (ZDZKKF02)
* **Corresponding author**
Conflicts of interest: Authors all declare no conflicts of interest.
Ethical approval: not required.
Informed consent: not required.
Cite this article: Sun RX, Zhao XL. Research advances in integrating multi-omics and artificial intelligence for the diagnosis and treatment of hematological malignancies[J]. Zaisheng Yixue Yanjiu, 2026, 1(3): 192-203. doi: 10.3969/j.issn.2097-7948.2026.01.010.

particularly machine learning and deep learning, have demonstrated substantial advantages in handling high-dimensional, multi-modal biomedical data, performing automated feature extraction, and constructing predictive models. They are evolving from auxiliary analytic tools into core approaches that drive diagnosis, risk stratification, and prediction of treatment response. This review outlines the characteristics of multi-omics data and the fundamental principles of AI methodologies, systematically summarizes the advances in their integrated application to molecular classification, relapse risk assessment and treatment response prediction in hematological malignancies, and analyses the existing limitations and key challenges, thereby providing new methodological perspectives and research directions for refining risk stratification systems and promoting precision diagnosis and treatment of hematological diseases.

Key words: Hematological malignancies; Multi-omics; Artificial intelligence; Precision medicine; Prognostic prediction

血液肿瘤是起源于造血系统的恶性克隆性疾病，其发生发展伴随正常造血功能紊乱和破坏，主要包括急、慢性白血病、各型淋巴瘤和多发性骨髓瘤（multiple myeloma，MM）等[1]。研究[1-3]表明血液恶性肿瘤占全部新发恶性肿瘤的 6% ~7%，死亡约占全部肿瘤死亡的 7%，不同地区和经济水平间存在显著差异。血液肿瘤具有起病隐匿、分子异质性高、疗效差异大的特点，传统依赖形态学和少数分子标志物的分层难以全面反映潜在生物学差异，对复发风险和个体化治疗获益的预测能力有限[1,3]。在最新版世界卫生组织（World Health Organization，WHO）与国际共识分类（International Consensus Classification，ICC）中，多种血液肿瘤已将特定驱动基因变异、染色体重排和表观遗传异常纳入诊断及风险分层核心依据，更凸显深度解析多维分子特征、构建精准诊断和预后工具的紧迫需求[4-7]。

人工智能（artificial intelligence，AI），尤其是机器学习（machine learning，ML）和深度学习（deep learning，DL），在近十余年伴随算力提升和大数据积累快速发展，已广泛应用于医学影像、信号及结构化/非结构化临床数据分析[8-11]。在血液学领域，AI 已被用于外周血涂片和骨髓形态学图像自动识别、多参数流式细胞术数据的群体划分及异常克隆检测、基因组/转录组特征筛选，以及电子病历风险预测等多个环节[5,8-9,11]。基于 AI 的诊断和预后模型已在急性白血病、骨髓增生异常综合征、淋巴瘤、骨髓瘤及遗传性贫血等疾病中显示出较高的可靠性和良好的准确性[5,8-9]。AI 正从辅助分析工具向临床决策支持系统演变，但要真正纳入规范化诊疗路径，仍需高质量数据集、透明可解释模型及多中心前瞻性验证的持续支撑[5,10-11]。

多组学是指在同一疾病或样本体系内综合获取基因组、转录组、表观遗传组、蛋白质组、代谢组及微生物组等多层信息并进行联合分析。相较单一标志物或单一通路研究，多组学可以从不同分子层面全面刻画肿瘤发生发展过程，更系统地揭示克隆演化、细胞异质性及其与微环境的互作[4-7]。单细胞和空间多组学的兴起，使研究者能够在单细胞分辨率和空间维度解析白血病克隆结构、免疫微环境细胞组成及细胞间互作，推进以分子分层为核心的血液肿瘤新分类体系[4,6]。然而，多组学数据体量巨大、结构复杂，传统统计方法在处理高维稀疏数据、非线性关联和跨平台批次效应时存在明显局限，这为 AI 介入并发挥优势提供了重要契机[4-5]。

多组学与 AI 的整合，是利用 ML/DL 模型对不同层级、不同来源的高维生物医学数据进行联合建模，以实现更精确的疾病表型刻画和更可靠的临床结局预测。在血液肿瘤的分子分型、复发风险评估和药物敏感性预测中，多组学联合 AI 模型总体优于单一组学或仅基于临床变量的模型[5]。将多组学大数据的深度与 AI 的广度相结合，是实现血液病精准诊疗的关键路径，不仅提升了诊断的精度和效率，更为实现血液疾病的个性化、动态化精准治疗开辟了新途径。

1　血液肿瘤的多组学解析

随着高通量测序和计算技术的迅速发展，血液肿瘤的研究已从单一维度的基因突变谱，扩展到覆盖基因组、表观遗传组、转录组、蛋白质组、代谢组以及空间组学的多维整合。多组学策略不仅有助于更精细地刻画克隆演化和耐药机制，还为传统分型与风险分层提供了补充，并为 AI 模型提供了高维度、结构化的数据基础[12]。

1.1 基因组与表观遗传组

在急性髓系白血病(acute myeloid leukemia, AML)中,大规模全基因组及甲基化测序推动了驱动事件与风险分层体系的持续更新。一项研究[12]整合 11 个队列共 3314 例急性白血病样本的 DNA 甲基化数据,构建了急性白血病甲基化组图谱(acute leukemia methylome atlas, ALMA),利用 ML 实现了对 WHO 2022 年第 5 版亚型的精确预测,并提出基于 38 个 CpG 位点的预后评分,其 5 年生存预测优于传统细胞遗传学和基因突变分层。

表观基因组在免疫治疗耐药中的作用也逐渐被重视。Ma 等[13]对接受 GPRC5D 靶向嵌合抗原受体 T(chimeric antigen receptor T, CAR-T)细胞治疗后复发的 MM 标本开展全基因组与全基因组亚硫酸氢盐测序,发现除少数病例存在 *GPRC5D* 基因双等位缺失外,更多患者 *GPRC5D* 下调来源于启动子及增强子区域的高度甲基化,且去甲基化药物可在体外恢复其表达。由此可见,血液肿瘤免疫逃逸不再局限于基因缺失等不可逆事件,而呈现出可逆的表观沉默,为通过去甲基化药物逆转免疫治疗耐药提供了新的生物学依据[14]。基因组与表观组数据通常具有极高的维度和稀疏性,且存在复杂的非线性互作。因此,适合采用支持向量机(support vector machine, SVM)或深度神经网络(deep neural networks, DNN)等能够处理高维稀疏特征的 AI 模型,通过降维和特征选择来识别关键的驱动变异及甲基化标记。

1.2 转录组

在转录组层面,单细胞 RNA 测序(single-cell RNA sequencing, scRNA-seq)结合表面蛋白、T 细胞受体(T cell receptor, TCR)/B 细胞受体(B cell receptor, BCR)以及染色质可及性等多组学数据,显著深化了对血液肿瘤克隆异质性和免疫微环境重塑的理解。Peretz 等[15]对混合表型急性白血病(mixed phenotype acute leukemia, MPAL)开展单细胞多组学分析,发现部分免疫表型模糊的亚群在转录组和染色质可及性上呈现高度干性特征,提示单靠形态及有限免疫表型可能低估干细胞样克隆的存在。在慢性髓系白血病(chronic myelogenous leukemia, CML)中,Warfvinge 等[16]将单细胞转录组、变异检测和药物敏感性测定整合,揭示了同一患者内部存在对酪氨酸激酶抑制剂(tyrosine kinase inhibitor, TKI)反应截然不同的克隆谱系:其中 BCR-ABL1 低表达、炎症通路上调的亚群在体外及随访中均表现为耐药或早期复发倾向。这些研究表明,多组学单细胞分析不仅能更精细地描绘克隆演化过程,并能与临床疗效相衔接,推动动态风险分层模型的建立。

单细胞转录组数据虽然分辨率高,但常伴随“丢弃事件(dropout events)”和高噪声。自编码器和生成对抗网络(generative adversarial network, GAN)等 DL 架构在数据降噪、缺失值填补及非线性降维方面表现优异,能从复杂的基因表达矩阵中有效提取细胞状态特征。

1.3 空间组学

传统测序方法往往忽略细胞在骨髓或淋巴结中的空间位置,而空间转录组及蛋白组技术则弥补了这一缺口。Steen 等[17-18]通过整合单细胞转录组、DNA 甲基化和空间转录组,对弥漫大 B 细胞淋巴瘤(diffuse large B-cell lymphoma, DLBCL)构建了“肿瘤细胞状态-生态位”图谱,发现特定 BCR 激活状态和代谢重编程的肿瘤细胞,常与髓样或免疫抑制性肿瘤相关巨噬细胞空间共定位,提示代谢-免疫相互作用在疾病进展中的关键作用。在 AML 中,Gui 等[19]利用单细胞空间转录组追踪接受免疫检查点抑制剂联合去甲基化药物治疗前后的骨髓标本,发现残留白血病细胞更倾向聚集于免疫效应细胞稀少、趋化因子低表达的局灶区域,并伴随成纤维细胞和内皮细胞转录谱重塑,为理解免疫治疗失败时“空间庇护所”的形成机制提供了直接证据。

空间组学数据的核心在于保留了细胞的物理位置信息。图神经网络(graph neural networks, GNN)能够将细胞视为节点,将空间邻域视为边,天然适配此类结构数据,从而有效建模肿瘤细胞与其微环境之间的空间通讯网络。

1.4 蛋白质组与代谢组

与基因和 RNA 水平相比,蛋白质及代谢物更直接反映信号通路活性和细胞功能状态。MM 的蛋白质组-基因组整合研究已鉴定出多个与药物耐受相关的信号轴,并通过磷酸化谱描绘出潜在的可药物化靶点网络[17-18]。在 AML 中,Nachmias 等[20]指出,白血病干细胞高度依赖线粒体氧化磷酸化

(oxidative phosphorylation, OXPHOS),并可通过与间充质干细胞之间的线粒体转运获得代谢优势。针对脂代谢和氨基酸代谢的研究工作进一步强调,特定代谢表型往往与独特的基因突变背景及肿瘤微环境重塑相耦联,为联合代谢抑制剂与现有靶向或免疫治疗提供了理论基础[20-22]。

蛋白质组与代谢组多为连续变量且缺失机制复杂,常需进行缺失值处理与批次校正。集成学习策略及基于知识图谱的 AI 方法,有助于将这些功能层面的表型数据与上游基因调控网络相关联,从而构建更完整的生物学信号通路。

2　血液肿瘤研究中 AI 与多组学的整合策略

AI 通常被界定为一类使计算机系统能够执行原本需要 AI 才能完成的任务的理论与技术集合,作为计算机科学的前沿分支,AI 正经历由传统 ML 向 DL 乃至生成式 AI 的快速演变[23]。自 2024 年以来,以Transformer 架构为核心的基座模型和大语言模型(large language models, LLMs)的兴起,为医学 AI 带来了革命性突破[24]。此类模型能够从高维、复杂且异质性强的组学与临床数据中识别潜在特征,为诊断、预后与治疗决策提供模型支持[9],形成覆盖图像诊断、基因组变异解析、转录组特征识别以及治疗反应预测的系统化应用框架[5, 25]。在诊断领域,AI 相关算法被用于自动识别外周血细胞与骨髓形态学特征、辅助流式细胞术免疫分型以及解释高通量测序结果[9, 26]。在预后评估方面,越来越多的研究尝试构建整合临床、生物标志物以及多层组学数据的高维风险模型,以改善复发预测、生存评估及药物敏感性预测的准确性[5]。随着多模态神经网络和医学领域预训练模型的出现,AI 具备处理纵向、多维度患者特征的能力,为血液疾病的个体化管理提供了新的技术路径。

2.1　多组学整合策略与数据预处理

多组学数据与 AI 的整合旨在通过系统化建模全面呈现血液疾病的分子图谱,其分析流程通常包括数据预处理、特征构建、组学集成策略选择以及模型训练与验证等关键步骤[27]。由于不同组学平台在测序深度、噪声结构和缺失机制上存在显著差异,高质量的前期处理是保证整合分析可靠性的前提。随后,研究者需依据生物学先验或算法驱动策略进行特征提取与降维,以构建信息密度高且具有稳定泛化能力的特征空间。在整合策略方面,以 Picard 等[27]提出的早期、混合、中间、后期与层级集成框架为代表的多组学整合方法,已成为当前 AI 建模的重要理论基础(表 1)。

高质量的数据预处理是 AI 建模的基石。不同组学的噪声结构与缺失机制差异显著,需要在建模前建立统一的质控与标准化流程,以降低技术噪声对 AI 模型训练的干扰(表 2)。近年来,多视图学习、GNN 以及深度生成模型逐渐成为解析组学间非线性关联与功能网络结构的重要工具,为解决传统线性方法难以捕获的跨组学依赖关系提供了新思路[28-29]。在模型训练与验证阶段,多采用随机森林、XGBoost、深度神经网络或图卷积网络等算法,并辅以交叉验证、外部独立队列验证及决策曲线分析综合评估模型的稳定性、预测效能与临床应用价值。

表 1　不同多组学整合策略的技术对比

Table 1　Comparison of multi-omics integration strategies

整合策略	融合层面	适用场景	代表性方法
早期整合	特征层面	样本配对完整;各组学数据尺度可比	特征拼接+Elastic Net/RF/XGBoost/DNN
混合整合	特征层面(单组学降维后)	组学异质性强;需要数据降维	单组学 PCA/自编码器/特征选择+ML
中间整合	共享潜在表示空间	存在共同驱动因素;允许部分组学缺失	MOFA+; NMF/JIVE/iCluster
后期整合	模型层面	异质性强;组学或样本存在部分重叠	堆叠集成或加权集成
层级整合	网络/通路层面	结构或先验知识驱动的生物学机制	SNF;图模型或通路模型

RF:随机森林;DNN:深度神经网络;PCA:主成分分析;NMF:非负矩阵分解;ML:机器学习;SNF:相似性网络融合。

RF: random forest; DNN: deep neural network; PCA: principal component analysis; NMF: non-negative matrix factorization; ML: machine learning; SNF: similarity network fusion.

2.2 AI 模型的可解释性技术在血液肿瘤中的应用

AI 模型的决策透明性不足一直是制约其广泛转化的关键障碍。注意力机制等可解释性方法被广泛用于阐明模型决策依据,以提升算法在临床场景中的可接受性和可信度[5, 26]。研究者在血液肿瘤领域引入了多种后置解释(post-hoc)技术,如 SHAP(SHapley Additive exPlanations)和 LIME(Local Interpretable Model-agnostic Explanations)等。在基于大规模 AML 数据构建的药物敏感性预测模型中,可解释性分析揭示特定基因或通路与药物反应密切相关,使 ML 模型的预测结果更具生物学合理性,并为临床理解关键分子驱动因素提供依据[30]。在淋巴瘤数字病理图像分类中,结合图像版 LIME 和 Grad-CAM 等可解释 AI(explainable AI, XAI)工具,可生成热图标注模型在分类过程中关注的组织学区域,有助于病理医生直观校验 AI 决策与组织特征之间的关联[31]。这类可解释性技术通过揭示特征贡献与视觉关注区域,大大提升了复杂 AI 模型的透明度与临床可用性,为 AI 在血液肿瘤诊疗中的转化提供了重要支撑。目前,多组学与 AI 整合技术已从初步的多数据源拼接发展为结合深度表征学习、网络结构建模与可解释分析的系统化框架,为实现血液疾病精准诊疗提供了坚实的数据与方法学基础(图 1)。

表 2 多组学数据标准化预处理与质控流程

Table 2 Standardized preprocessing and quality control workflows for multi-omics data

数据类型	关键质控指标	典型工具
基因组学	测序深度、比对率、重复序列;变异质量/深度	BWA/Samtools/Picard;GATK;ANNOVAR
转录组学	比对率,rRNA 比例;离群样本基因检出数/UMI 计数;线粒体比例;双细胞率	STAR /HISAT2;DESeq2/Seurat/SCTransform;ComBat-Seq/Harmony
表观组学	探针/峰值覆盖度;样本质控	Minfi/ChAMP;MACS2/ArchR/Signac
蛋白质学/代谢组学	信号漂移;变异系数;缺失率	QC-RLSC;MaxQuant/Proteome Discoverer;XCMS
空间组学	组织覆盖/捕获点质控;空白捕获点图像质量;空间配准	Seurat Spatial;Squidpy;Giotto;BayesSpace

rRNA:核糖体 RNA;UMI:唯一分子标识。

rRNA: ribosomal RNA; UMI: unique molecular identifier.

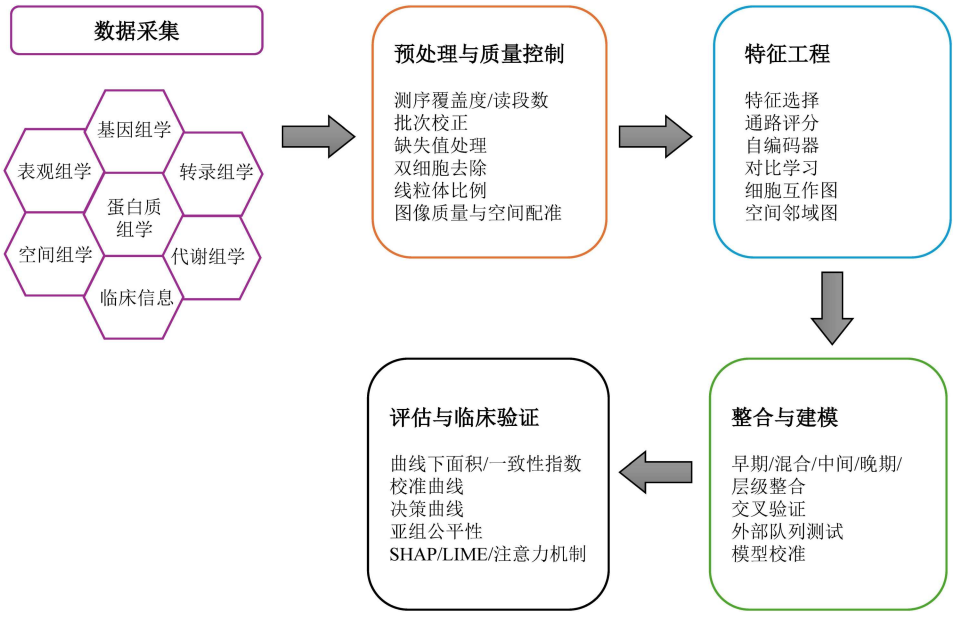


图 1 多组学与 AI 整合的技术路线图。AI:人工智能。

Figure 1 Technical roadmap for multi-omics integration with AI. AI: artificial intelligence.

3　多组学与 AI 整合在血液肿瘤诊疗中的应用

随着组学技术和计算方法的迅速发展，血液疾病研究正逐步从以单一组学结合传统统计分析为主的模式，转向以 AI 为核心、整合多组学与临床多模态信息的系统化研究框架。DL 和多模态建模使大规模基因组、转录组、蛋白质组、代谢组以及影像与临床数据能够在同一模型中联合训练，为肿瘤分型、靶点发现、疗效预测与随访管理提供新的技术路径[32-34]。在白血病、淋巴瘤和 MM 等血液恶性肿瘤中，多组学与 AI 的整合研究已从概念验证逐步走向接近临床转化的阶段，成为推动精准诊疗和预后评估的重要支撑力量（图 2）。

3.1　多组学与 AI 协同优化分子分型与精准诊断

在分子层面，多组学整合与 AI 的结合正在重塑传统分类与风险分层体系。一方面，转录组与表观组学的联合建模能在最小化人工特征工程的前提下高效构建诊断与分型工具。Warnat-Herresthal 等[35]整合 12 029 份外周血转录组样本，对多种 ML 算法进行系统比较，在跨队列情境下仍具良好泛化能力的 AML 预测模型，并定量评估了特征维度、批次效应及算法选择对诊断性能的影响，为大规模转录组-ML 诊断提供了方法学范式。另一方面，前文提及的 ALMA研究则以多队列甲基化数据的统一预处理和降维建模为基础，在数据驱动的疾病空间中进一步训练急性白血病亚型分类和预后模型，从方法学层面展示了以全基因组甲基化结合 ML 实现 WHO 2022 年第 5 版分类与风险分层的可行性，并与其在表观遗传调控和预后分层方面的生物学发现形成了相互印证[36]。

除单一表观组学外，将转录组、miRNA、拷贝数变异等多维数据通过多组学聚类与监督学习共同建模，可揭示传统分型无法覆盖的生物学亚群。Wang 等[37]基于 126 例 AML 患者的转录组、miRNA 及拷贝数变异等多组学数据，采用多组学聚类和 ML 构建分层模型，将患者划分为 4 个分子亚群，在基因突变谱、免疫微环境特征、药物敏感性及生存结局方面存在显著差异。该模型在外部 GEO 队列中获得了良好的稳健性，为传统法美英协作组（French-American-British，FAB）/WHO 分类及欧洲白血病网络（European Leukemia Network，ELN）风险分层提供了有力补充。

在淋巴瘤领域，多组学与 AI 整合的诊断策略也取得了快速进展。整合基因突变、转录组、蛋白表达及影像特征的综合模型，可在传统 DLBCL 分型基础上进一步细化分子亚型，并关联各亚型的信号通路激活模式与免疫微环境特征[38]。Zhang 等[39]进一步将血浆循环游离 DNA（circulating free DNA，cfDNA）、miRNA 和代谢物联合建模，构建了用于 DLBCL 早期诊断和化疗反应预测的非侵入性多组学联合 AI 模型，提出了以液体活检为核心的整体诊断与监测策略。

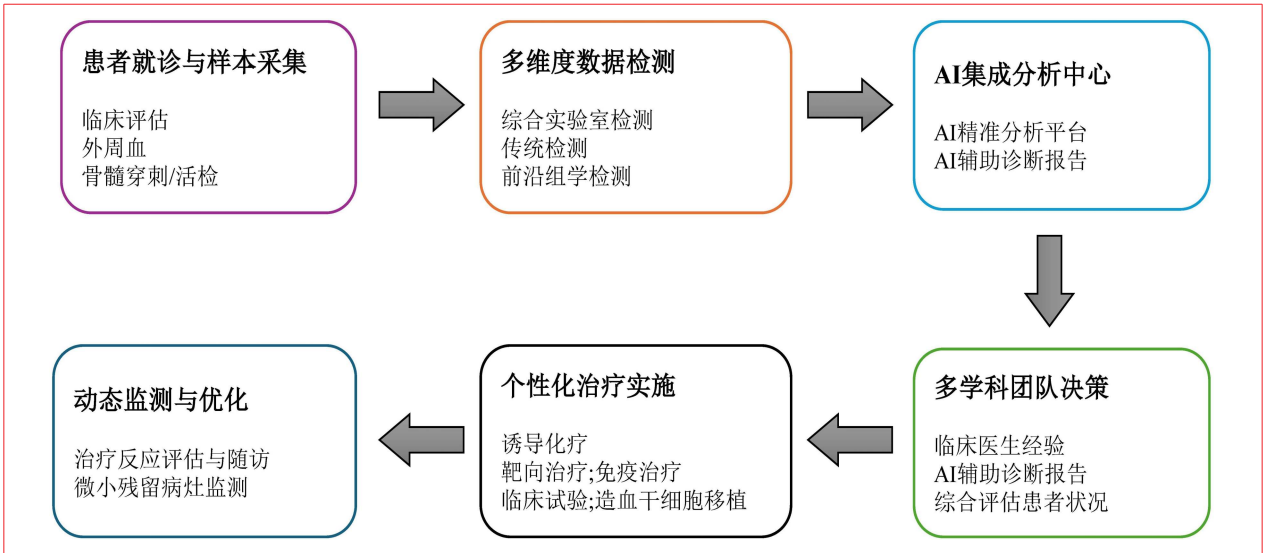


图 2　血液肿瘤诊疗中 AI 赋能的临床应用流程图

Figure 2　AI-enabled clinical workflow for the diagnosis and management of hematologic malignancies

MM 的蛋白质组与基因组整合研究则提供了另一种思路。Ramberger 等[40]报道的蛋白基因组图谱揭示了不同染色体异常与蛋白表达模式之间的系统性关系，并辅以 ML 挖掘疾病亚型及潜在可药物化靶点网络。在此基础上，Chang 等[41]提出的 MyeVAE 模型将转录组、基因突变和临床特征共同输入多模态变分自编码器，再结合生存子网络实现风险预测，C-index 明显优于国际分期系统（International Staging System，ISS）/修订版 ISS（Revised ISS，R-ISS）等传统分层系统。这些工作表明，多组学与 AI 的联合不仅可在单病种内部建立更精细的分子分型体系，也为跨疾病亚型、跨平台数据的统一表示提供了技术基础。

3.2 多组学与 AI 联合解析治疗反应与耐药机制

在治疗反应与耐药机制研究中，多组学与 AI 的整合同样展现出独特优势。针对 CAR-T 细胞治疗，多项研究采用单细胞多组学技术联合 ML 描绘了 CAR-T 细胞及其免疫微环境的动态变化。Guerrero-Murillo 等[42]在复发/难治 B 细胞恶性肿瘤患者的 CD19 CAR-T 治疗中，采用单细胞转录组、表观组和表面蛋白等多组学联合分析，通过无监督聚类和图神经网络解析 CAR^+ 与 CAR-T 细胞状态，识别出具有持续增殖和长期存活特征的 CAR-T 细胞亚群，并揭示其转录调控网络与早期耗竭表型显著不同，为优化生产工艺和质量放行指标提供了可量化的功能标志物。将此类细胞状态特征与临床结局和药代动力学数据进一步联结，可构建由细胞功能谱推断疗效的预测模型，为实时响应评估和剂量调整提供决策基础。

耐药和复发仍然是血液肿瘤治疗失败及长期生存受限的关键因素。结合单细胞测序与 AI 分析可在克隆进化层面追踪耐药亚群的轨迹。对 AML 患者治疗前后样本进行单细胞 DNA 分型，并利用算法重建克隆进化树，可以识别在治疗压力下选择性扩增的微小耐药亚克隆（如伴 *TP53* 突变或 RAS 通路持续激活的细胞），这些亚克隆常在形态学复发前数月即已出现，为提前识别高危人群和调整治疗策略提供了宝贵的时间窗口[43]。

在 MM 中，基于 AI 的多组学整合分析提示，达雷妥尤单抗（Daratumumab）耐药不仅与 CD38 抗原丢失有关，还涉及表观遗传修饰酶 KDM6A 的功能缺失。KDM6A 缺失通过上调 H3K27me3 水平导致染色质重塑，进而下调 CD38 和 CD48 的表达，削弱了自然杀伤细胞介导的抗体依赖性细胞介导的细胞毒性作用（antibody-dependent cell-mediated cytotoxicity，ADCC）。基于此机制，联合使用 EZH2 抑制剂能够降低 H3K27me3 水平，可恢复 CD38/CD48 表达，从而为联合表观调控药物逆转耐药提供了重要的机制依据和潜在治疗策略[44]。

3.3 多组学与 AI 驱动的风险分层与预后预测

多组学与 AI 整合在血液肿瘤风险分层和预后预测中的优势正逐步显现。AML 预后具有异质性，而现行 ELN 风险分层在中间风险组及基因组结构复杂等部分亚群中的预后差异识别能力仍然不足。Kosvyra 等[45]利用基因表达、突变谱和临床变量构建多层网络特征，并基于 ML 与网络分析建立生存预测模型，结果显示多组学网络指标可显著提升 2 年总生存预测的准确性，且在外部队列中保持良好的稳健性。Zhang 等[46]提出的 AML 生存预测（AML survival prediction，AMLSP）方法通过对多模态特征进行解耦与重融合，在多个人群背景和检测平台的独立 AML 数据集中取得 0.7 ~ 0.8 的 C-index，展示了在高度异质性疾病中利用多组学与 AI 整合提升预后预测性能的可行性。

在 DLBCL 中，多组学与 AI 的联合建模亦显著改善了风险分层精度与临床可解释性。Wang 等[38]将正电子发射型计算机断层显像（positron emission computed tomography，PET）/计算机断层扫描（computed tomography，CT）影像深层特征与转录组、基因突变及临床变量综合建模的深度学习框架，较传统国际预后指数（international prognostic index，IPI）等评分系统在无进展生存期和总生存期预测方面均表现更优，并可用于区分对标准 R-CHOP 方案不敏感的高危人群；Zhang 等[39]开发的基于血浆多组学的非侵入性 AI 模型还可同步评估疾病状态与化疗敏感性，有助于在治疗起始阶段即进行个体化方案的选择。

MM 领域中，除 MyeVAE 模型外，Xie 等[47]构建的单核细胞驱动多组学 ML 预后模型，将肿瘤转录组特征、免疫浸润状态以及外周单核细胞相关分子整合

于统一框架中，能够有效识别高危患者并揭示免疫微环境在疾病进展中的核心作用；Feng 等[48]则利用 PANoptosis 相关基因，结合多队列转录组和临床数据，通过 *LASSO* 回归和随机森林等方法构建 PANoptosis-相关预后评分，并在多中心外部队列中验证其稳健性，显示该模型在预后评估和免疫治疗策略选择中的潜在应用价值。表 3 对多组学与 AI 整合的相关研究进行了总结。以上研究共同表明，多组学与 AI 模型在多种血液肿瘤的传统预后系统基础上实现显著的性能提升，并具有更高临床应用价值。

4　血液疾病诊疗中多组学与 AI 的局限性

尽管多组学与 AI 的结合在血液疾病诊疗方面展现出巨大潜力并取得诸多成果，但目前在实践应用中仍存在诸多局限性和挑战。在数据层面存在多中心标准化与外部验证不足。模型性能高度依赖于数据的质量和数量，现有研究总体仍停留在单中心、小样本、回顾性探索阶段，疾病谱与人群构成高度不均衡，难以支撑高复杂度模型的稳健训练和跨队列外推，真正实现多层组学深度整合并开展外部验证的工作仍然有限[5]。解决路径应强调“统一采集-质控-标注-评价”标准与多中心前瞻性验证。以血液学数字形态学为例，Scopio X100 的多中心验证研究[49]提供了从多中心证据链到临床可用的可复制范式，可为血液肿瘤 AI 转化时的数据治理参照。

在技术层面，稳健性与系统集成仍是多组学 AI 临床落地的关键瓶颈。不同组学数据在检测深度、噪声结构、动态范围及缺失模式方面存在系统差异，批次效应和平台差异进一步加剧了跨中心整合的难度。在此背景下，即便是 DL 等非线性模型，也容易在训练集中获得较高性能而在独立队列上明显退化[33]。例如骨髓细胞形态分析模型 DeepHeme 在外部中心验证时，平均 F1 由 0.89 降至 0.85，且血小板聚集等类别的性能下降与制片伪影差异相关[50]。此外，目前尚缺乏围绕血液疾病多组学构建的统一数据标准化、质控流程和基准数据集，不同研究间的分析流程和评价指标差异较大，削弱了结果的可比性和可复现性。为了让模型真正能落到临床，技术上更可行的做法是把 AI 做成人机协同的决策支持系统（decision support system，DSS），模型给出可解释证据、可复核的中间结果并接入 LIS/EMR，关键判断仍由专业人员确认、可追溯审计[26]。

表 3　血液肿瘤中多组学与 AI 整合研究进展

Table 3　Advances in multi-omics-AI integration in hematologic malignancies

应用	疾病类型	整合策略	关键发现
分子分型与精准诊断	AML	DNA 甲基化+机器学习分类器	实现 WHO 2022 年第 5 版亚型预测；发现集准确率 96.3%，验证集准确率 90.1%
	DLBCL	cfDNA 多组学+集成机器学习	实现非侵入性早期诊断；曲线下面积达 0.993；在 98% 特异性下敏感性为 95%；早期敏感性在 99% 特异性下为 91%
	MM	多组学+变分自编码器	风险分层；验证集总生存期 C-index 约为 0.70
治疗反应与耐药机制	B 细胞淋巴瘤	单细胞多组学+图神经网络	解析 CAR-T 相关细胞状态/特征；支持细胞生产与疗效监测
	AML	单细胞 DNA+克隆演化模型	基于 scDNA-seq（123 例 AML）解析克隆结构；将演化模式与耐药/复发相关联
	MM	表观组学+功能验证	阐明 KDM6A 调控 CD38/CD48 机制；为免疫/抗体反应调节提供机制基础
风险分层与预后预测	DLBCL	PET/CT 多模态深度特征+自动化机器学习+*Cox*	生存预测；无进展生存期 C-index 为 0.784/0.739，总生存期 C-index 为 0.831/0.782
	AML	多模态特征+网络建模	指出异质性背景下的风险建模需要严格的外部验证
	MM	免疫表型+机器学习（*LASSO*/随机森林）	建立机器学习风险评分；总生存期 C-index 为 0.667

AML：急性髓系白血病；DLBCL：弥漫大 B 细胞淋巴瘤；MM：多发性骨髓瘤；cfDNA：循环游离 DNA；WHO：世界卫生组织；PET：正电子发射型计算机断层显像；CT：计算机断层扫描；CAR-T：嵌合抗原受体 T。

AML：acute myeloid leukemia；DLBCL：diffuse large B-cell lymphoma；MM：multiple myeloma；cfDNA：circulating free DNA；WHO：World Health Organization；PET：positron emission computed tomography；CT：computed tomography；CAR-T：chimeric antigen receptor T.

在监管层面，伦理和法律方面的挑战不容忽视。AI 在医疗中的应用可能对传统医患关系产生影响，例如过度依赖 AI 可能削弱医师的决策主动性。隐私与数据安全问题同样制约了大规模多组学队列的构建与共享[8]。集中存储基因组和临床数据易引发隐私泄露隐患，而跨机构数据流动又受到严格监管。此外，算法本身可能存在偏倚，如果训练数据不平衡，AI 决策可能放大某些群体在医疗服务获得上的不公平。因此，制定明确的法规和行业标准，确保 AI 应用的安全性、公平性和透明度，是 AI 医疗推广过程中必须面对的问题。

在临床层面，AI 模型的可解释性不足也限制了其在临床中的信任度和应用价值。多组学与 AI 模型往往采用多层神经网络或集成学习等复杂结构，特征交互高度非线性，导致决策过程难以用传统统计学思路加以解释。这种缺乏透明度的黑箱模型在医疗决策中难以获得医生和患者的充分信任，阻碍了 AI 系统在临床实践中的广泛应用。将医学领域的先验知识融入模型或借助大语言模型生成对结果的解释，是提高模型可解释性的有益尝试。提升 AI 模型的透明度和可解释性，有助于医生更好地理解和校验模型输出，从而在临床中更愿意采纳 AI 建议。

在未来的研究和推广中，我们需要针对上述问题不断改进技术和策略，以充分释放 AI 在血液疾病精准诊疗中的潜能。

5 挑战与展望

多组学与 AI 技术的协同应用在血液疾病诊疗与预后预测中具有革命性的潜力，但要从概念验证走向临床落地，未来研究还需要聚焦更多方向：(1) 引入生成式 AI 以缓解小样本与缺失模态问题，以 GAN 与 Transformer 为代表的生成式模型能够学习高维组学分布，用于合成数据增强、缺失补全与“虚拟队列”构建；(2) 构建高质量、多中心基准数据库与统一质控标准，类似 TCGA 的血液肿瘤专用资源，应同时覆盖多组学、治疗方案与结局随访，并以统一的采样、测序/质控、标注与评测指标支撑跨队列泛化；(3) 推进“多组学-AI+再生医学”的场景化融合：围绕造血干细胞移植等再生医学关键环节，将移植前后免疫微环境与多组学信息纳入建模，发展可解释的动态风险预警（如植入失败、移植物抗宿主病、复发风险），并与临床决策系统联动形成可执行的干预策略。通过解决当前面临的挑战，推动技术融合与临床验证，我们有望在不久的将来见证 AI 赋能下的血液疾病精准诊疗新局面的到来，为患者带来更为长远的生存和更高的生活质量。从高分辨的生物学解析到智能化的临床决策，多组学与 AI 的整合有望从概念验证阶段迈向真正嵌入血液疾病诊疗流程的临床决策支持工具。

参考文献

1 Zhang N, Wu J, Wang Q, *et al*. Global burden of hematologic malignancies and evolution patterns over the past 30 years [J]. Blood Cancer J, 2023, 13(1): 82. doi: 10.1038/s41408-023-00853-3.

2 Sun K, Wu H, Zhu Q, *et al*. Global landscape and trends in lifetime risks of haematologic malignancies in 185 countries: population-based estimates from GLOBOCAN 2022 [J]. EClinicalMedicine, 2025, 83: 103193. doi: 10.1016/j.eclinm.2025.103193.

3 Bray F, Laversanne M, Sung H, *et al*. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries [J]. CA Cancer J Clin, 2024, 74(3): 229-263. doi: 10.3322/caac.21834.

4 Pirouzbakht M, Hamzeh S, Soleimani Samarkhazan H. Beyond single biomarkers: multi-omics strategies to predict immunotherapy outcomes in blood cancers [J]. Clin Exper Med, 2025, 25(1): 1-18. doi: 10.1007/s10238-025-01902-w.

5 Alhumrani SQ, Ball GR, El-Sherif AA, *et al*. Machine learning for multi-omics characterization of blood cancers: a systematic review [J]. Cells, 2025, 14(17): 1385. doi: 10.3390/cells14171385.

6 Dong M, Wang L, Hu N, *et al*. Integration of multi-omics approaches in exploring intra-tumoral heterogeneity [J]. Cancer Cell Int, 2025, 25(1): 317. doi: 10.1186/s12935-025-03944-2.

7 Liu X, Wei K, Lin E, *et al*. Multi-omics advances in under-

standing cancer drug resistance[J]. Biomed Technol, 2025, 12: 100119. doi: 10.1016/j.bmt.2025.100119.

8 Radakovich N, Nagy M, Nazha A. Artificial intelligence in hematology: current challenges and opportunities[J]. Curr Hematol Malig Rep, 2020, 15(3): 203-210. doi: 10.1007/s11899-020-00575-4.

9 El Alaoui Y, Elomri A, Qaraqe M, *et al*. A review of artificial intelligence applications in hematology management: current practices and future prospects[J]. J Med Internet Res, 2022, 24(7): e36490. doi: 10.2196/36490.

10 Haug CJ, Drazen JM. Artificial intelligence and machine learning in clinical medicine, 2023[J]. N Engl J Med, 2023, 388(13): 1201-1208. doi: 10.1056/NEJMra2302038.

11 Walter W, Haferlach C, Nadarajah N, *et al*. How artificial intelligence might disrupt diagnostics in hematology in the near future[J]. Oncogene, 2021, 40(25): 4271-4280. doi: 10.1038/s41388-021-01861-y.

12 Khan H, Vale C, Bhagat T, *et al*. Role of DNA methylation in the pathogenesis and treatment of myelodysplastic syndromes[J]. Semin Hematol, 2013, 50(1): 16-37. doi: 10.1053/j.seminhematol.2013.01.001.

13 Ma S, Xia J, Zhang M, *et al*. Genetic and epigenetic mechanisms of *GPRC5D* loss after anti-GPRC5D CAR T-cell therapy in multiple myeloma[J]. Blood, 2025, 146(2): 178-190. doi: 10.1182/blood.2024026622.

14 Lee H, Bahlis NJ. Silence of the myeloma clones: GPRC5D epigenetic regulation[J]. Blood, 2025, 146(2): 133-135. doi: 10.1182/blood.2025028865.

15 Peretz CA, Kennedy VE, Walia A, *et al*. Multiomic single cell sequencing identifies stemlike nature of mixed phenotype acute leukemia[J]. Nat Commun, 2024, 15(1): 8191. doi: 10.1038/s41467-024-52317-2.

16 Warfvinge R, Ulfsson LG, Dhapola P, *et al*. Single-cell multiomics analysis of chronic myeloid leukemia links cellular heterogeneity to therapy response[J]. Elife, 2024, 12: RP92074. doi: 10.7554/eLife.92074.

17 Steen CB, Luca BA, Esfahani MS, *et al*. The landscape of tumor cell states and ecosystems in diffuse large B cell lymphoma[J]. Cancer Cell, 2021, 39(10): 1422-1437. e10. doi: 10.1016/j.ccell.2021.08.011.

18 Dai L, Fan G, Xie T, *et al*. Single-cell and spatial transcriptomics reveal a high glycolysis B cell and tumor-associated macrophages cluster correlated with poor prognosis and exhausted immune microenvironment in diffuse large B-cell lymphoma[J]. Biomarker Res, 2024, 12(1): 58. doi: 10.1186/s40364-024-00605-w.

19 Gui G, Bingham MA, Herzog JR, *et al*. Single-cell spatial transcriptomics reveals immunotherapy-driven bone marrow niche remodeling in AML[J]. Sci Adv, 2025, 11(28): eadw4871. doi: 10.1126/sciadv.adw4871.

20 Nachmias B, Wolach O. Beating the STATs: targeting the metabolome in acute myeloid leukemia[J]. Haematologica, 2026, 111(2): 439-441. doi: 10.3324/haematol.2025.288349.

21 Pereira-Martins DA, Weinhauser I, Griessinger E, *et al*. High mtDNA content identifies oxidative phosphorylation-driven acute myeloid leukemias and represents a therapeutic vulnerability[J]. Signal Transduction Targeted Ther, 2025, 10(1): 222. doi: 10.1038/s41392-025-02303-x.

22 Dembitz V, James SC, Gallipoli P. Targeting lipid metabolism in acute myeloid leukemia: biological insights and therapeutic opportunities[J]. Leukemia, 2025, 39(8): 1814-1823. doi: 10.1038/s41375-025-02645-z.

23 Gignac GE, Szodorai ET. Defining intelligence: bridging the gap between human and artificial perspectives[J]. Intelligence, 2024, 104: 101832. doi: 10.1016/j.intell.2024.101832.

24 Liu Z, Zhao Z, Shi Z. Medical imaging foundation models: paradigm innovation in precision oncology[J]. Med J PUMCH, 2025, 16(4): 805-811. doi: 10.12290/xhyxzz.2025-0328.

25 Walter W, Pohlkamp C, Meggendorfer M, *et al*. Artificial intelligence in hematological diagnostics: game changer or gadget?[J]. Blood Rev, 2023, 58: 101019. doi: 10.1016/j.blre.2022.101019.

26 Baiao AR, Cai Z, Poulos RC, *et al*. A technical review of multi-omics data integration methods: from classical statistical to deep generative approaches[J]. Brief Bioinform, 2025, 26(4): bbaf355. doi: 10.1093/bib/bbaf355.

27 Picard M, Scott-Boyer MP, Bodein A, *et al*. Integration strategies of multi-omics data for machine learning analysis

〔J〕. Comput Struct Biotechnol J, 2021, 19: 3735-3746. doi: 10.1016/j.csbj.2021.06.030.

28 Sanches PHG, De Melo NC, Porcari AM, *et al.* Integrating molecular perspectives: strategies for comprehensive multi-omics integrative data analysis and machine learning applications in transcriptomics, proteomics, and metabolomics〔J〕. Biology, 2024, 13(11): 848. doi: 10.3390/biology13110848.

29 Fouché A, Zinovyev A. Omics data integration in computational biology viewed through the prism of machine learning paradigms〔J〕. Front Bioinform, 2023, 3: 1191961. doi: 10.3389/fbinf.2023.1191961.

30 Nasimian A, Al Ashiri L, Ahmed M, *et al.* A receptor tyrosine kinase inhibitor sensitivity prediction model identifies AXL dependency in leukemia〔J〕. Int J Mol Sci, 2023, 24(4): 3830. doi: 10.3390/ijms24043830.

31 Carreras J, Ikoma H, Kikuti YY, *et al.* Histological image classification between follicular lymphoma and reactive lymphoid tissue using deep learning and explainable artificial intelligence (XAI)〔J〕. Cancers, 2025, 17(15): 2428. doi: 10.3390/cancers17152428.

32 Marouf AA, Rokne JG, Alhajj R. Integrating multi-omics and medical imaging in artificial intelligence-based cancer research: an umbrella review of fusion strategies and applications〔J〕. Cancers, 2025, 17(22): 3638. doi: 10.3390/cancers17223638.

33 Zhang J, Che Y, Liu R, *et al.* Deep learning-driven multi-omics analysis: enhancing cancer diagnostics and therapeutics〔J〕. Brief Bioinform, 2025, 26(4): bbaf440. doi: 10.1093/bib/bbaf440.

34 Ballard JL, Wang Z, Li W, *et al.* Deep learning-based approaches for multi-omics data integration and analysis〔J〕. BioData Min, 2024, 17(1): 38. doi: 10.1186/s13040-024-00391-z.

35 Warnat-Herresthal S, Perrakis K, Taschler B, *et al.* Scalable prediction of acute myeloid leukemia using high-dimensional machine learning and blood transcriptomics〔J〕. Iscience, 2020, 23(1): 100780. doi: 10.1016/j.isci.2019.100780.

36 Marchi F, Shastri VM, Marrero RJ, *et al.* Epigenomic diagnosis and prognosis of acute myeloid leukemia〔J〕. Nat Commun, 2025, 16(1): 6961. doi: 10.1038/s41467-025-62005-4.

37 Wang T, Cui S, Lyu C, *et al.* Molecular precision medicine: multi-omics-based stratification model for acute myeloid leukemia〔J〕. Heliyon, 2024, 10(17): e36155. doi: 10.1016/j.heliyon.2024.e36155.

38 Wang Y, Wang X, Huang X, *et al.* Integration of deep learning in baseline PET/CT and multi-omics data for prognostic modeling in diffuse large B-cell lymphoma: the clinicalpet lymphplex model; proceedings of the 66th ASH Annual Meeting Abstracts, F, 2024〔C〕. American Society of Hematology Washington, DC. doi: 10.1182/blood-2024-206537.

39 Zhang W, Ye B, Song Y, *et al.* Integrating multi-omics features enables non-invasive early diagnosis and treatment response prediction of diffuse large B-cell lymphoma〔J〕. Clin Transl Med, 2025, 15(1): e70174. doi: 10.1002/ctm2.70174.

40 Ramberger E, Sapozhnikova V, Ng YLD, *et al.* The proteogenomic landscape of multiple myeloma reveals insights into disease biology and therapeutic opportunities〔J〕. Nat Cancer, 2024, 5(8): 1267-1284. doi: 10.1038/s43018-024-00784-3.

41 Chang JG, Chen J, Chew GL, *et al.* MyeVAE: a multi-modal variational autoencoder for risk profiling of newly diagnosed multiple myeloma〔J〕. BMC Artif Intell, 2025, 1(1): 1-17. doi: 10.1186/s44398-025-00009-2.

42 Guerrero-Murillo M, Rill-Hinarejos A, Trincado JL, *et al.* Integrative single-cell multi-omics of CD19-CARpos and CARneg T cells suggest drivers of immunotherapy response in B cell neoplasias〔J〕. Cell Rep Med, 2024, 5(11): 101803. doi: 10.1016/j.xcrm.2024.101803.

43 Morita K, Wang F, Jahn K, *et al.* Clonal evolution of acute myeloid leukemia revealed by high-throughput single-cell genomics〔J〕. Nat Commun, 2020, 11(1): 5327. doi: 10.1038/s41467-020-19119-8.

44 Liu J, Xing L, Li J, *et al.* Epigenetic regulation of CD38/CD48 by KDM6A mediates NK cell response in multiple myeloma〔J〕. Nat Commun, 2024, 15(1): 1367. doi: 10.1038/s41467-024-45561-z.

45 Kosvyra A, Karadimitris A, Papaioannou M, *et al.* Machine learning and integrative multi-omics network analysis for

survival prediction in acute myeloid leukemia〔J〕. Comput Biol Med, 2024, 178: 108735. doi: 10.1016/j.compbiomed.2024.108735.

46 Zhang S, Lin X, Yao Y, *et al*. Amlsp model for AML survival prediction; proceedings of the 67th ASH Annual Meeting Abstracts, F, 2025〔C〕. American Society of Hematology Washington, DC. doi: 10.1182/blood-2025-393.

47 Xie L, Gao M, Tan S, *et al*. A monocyte-driven prognostic model for multiple myeloma: multi-omics and machine learning insights〔J〕. Blood Lymphat Cancer, 2025, 15: 21-37. doi: 10.2147/BLCTT.S517354.

48 Feng Y, Wang S, Zhang J, *et al*. Comprehensive machine learning analysis of PANoptosis signatures in multiple myeloma identifies prognostic and immunotherapy biomarkers〔J〕. Sci Rep, 2025, 15(1): 22478. doi: 10.1038/s41598-025-06376-0.

49 Katz BZ, Feldman MD, Tessema M, *et al*. Evaluation of Scopio Labs X100 Full Field PBS: The first high-resolution full field viewing of peripheral blood specimens combined with artificial intelligence-based morphological analysis〔J〕. Int J Lab Hematol, 2021, 43(6): 1408-1416. doi: 10.1111/ijlh.13681.

50 Sun S, Yin Z, Van Cleave JG, *et al*. DeepHeme, a high-performance, generalizable deep ensemble for bone marrow morphometry and hematologic diagnosis〔J〕. Sci Transl Med, 2025, 17(802): eadq2162. doi: 10.1126/scitranslmed.adq2162.

（责任编辑：胡艳杰）