



综合多组学数据的肝细胞癌分层策略及进展*

王 猛 李晓琴** 高 斌

(北京工业大学环境与生命学部, 北京 100124)

摘要 肝细胞癌 (hepatocellular carcinoma, HCC) 是最常见和致命的肝脏恶性肿瘤。这种疾病的治疗一直受到其异质性的阻碍, 极大限制了其个性化治疗的进展。因此, 将高度异质的HCC分成具有相似特征的分子亚类对其临床治疗有着重要意义。随着高通量技术的不断发展, 多种组学数据的关联研究可以加深了解HCC发生背后的生物学机制, 也为HCC分层研究打开了新的思路。本文对当前HCC多组学分层策略及其相关研究进行了综述, 并总结了当前HCC亚型的多组学特征。

关键词 肝细胞癌, 多组学, 亚型识别

中图分类号 Q31, R735.7

DOI: 10.16476/j.pibb.2022.0397

肝细胞癌 (hepatocellular carcinoma, HCC) 是最常见的原发性肝脏恶性肿瘤, 其占比约90%。根据最新全球癌症数据统计, HCC已成为全球第六大最常诊断的癌症和第三大癌症死亡原因, 其发病率在全球范围内呈上升趋势^[1-2]。在中国, HCC是中国第四大常见癌症和第二大癌症相关死亡原因^[3], 严重威胁中国人民的生命和健康。流行病学调查显示, 慢性肝炎和肝硬化是全球范围内诱发HCC的主要风险因素^[4]。除此之外, 由代谢综合征、肥胖、2型糖尿病和非酒精性脂肪肝等风险因素诱发的HCC患病率正逐年增加, 并在未来可能成为全球HCC发生的主要原因^[5]。从分子角度, HCC是由多种基因组和表观基因组改变的累积引起。其中常见的包括 *TERT* 启动子、*AXIN1*、*TP53* 和 *CTNNB1* 的致癌突变, 染色体 1q、8q 的扩增, 8p、22p 的丢失等^[6]。这些发生遗传改变的基因往往与 Wnt- β -catenin、细胞周期控制、AKT-mTOR、MAPK 等通路有关^[7]。突变特征分析也表明, 一些诱发DNA突变的风险因素如饮酒、吸烟以及黄曲霉毒素B1暴露等, 也与已知基因的致癌突变相关^[8]。这似乎也表明, 肝脏虽然可以在体内起到解毒作用, 但这些有毒代谢物也可以诱导相关基因的驱动突变进而损害肝细胞基因组, 从而导致癌变。

目前用于HCC分型的系统主要基于肿瘤负荷, 在临床上, 巴塞罗那临床肝癌分期系统是迄今为止预测HCC预后和指导选择治疗干预措施最常用的模型^[9]。而中国则根据本国国情及实践积累, 依据患者体力活动状态、肝肿瘤及肝功能情况, 提出了中国肝癌分期系统, 共分为四期^[10]。但值得注意的是, 这些分期系统虽然对治疗方案的选择和预后评估至关重要, 但它们不足以描述影响预后和治疗反应的生物学和分子特征^[11]。因此十分必要开发一个HCC精确的分子分类系统。

随着高通量技术的不断发展和其成本的下降, 癌症基因组图谱 (The Cancer Genome Atlas, TCGA)、国际癌症基因组联盟 (International Cancer Genome Consortium, ICGC) 和肿瘤细胞系的百科全书 (Cancer Cell Line Encyclopedia, CCLE) 等国际合作项目已收集了同一癌症患者队列不同层次的组学数据。这为癌症分子分型提供重要依据, 并能够很好地反映不同亚型下癌症生物学背景的差异, 对癌症的治疗有着重要影响。单一组

* 国家自然科学基金 (61931013) 和国家重点研发计划 (2017YFC011104) 资助项目。

** 通讯联系人。

Tel: 15313254516, E-mail: lxq0811@bjut.edu.cn

收稿日期: 2022-08-22, 接受日期: 2022-10-07

学的研究虽然可以单向揭示肿瘤大量信息（如转录组数据可以描述癌症之间基因表达差异），但癌症与宿主之间相互作用、癌症内部分子之间相互作用以及不同组学之间关联需要多维方法来描绘。因此，多组学整合研究（即整合两个或多个组学数据，进行数据分析、可视化和解释）被认为是深入了解癌症病发机制和癌症异质性最有前景的工具^[12-13]。本文总结了当前HCC多组学分层策略和相关研究进展。

1 HCC的异质性

HCC是一种在病理和分子水平上高度异质性的疾病，这种异质性可能源于不同的风险因素、遗传事件、基因表达模式、激活的通路、免疫浸润程度或肿瘤间质变化，其大致可分为瘤内异质性和瘤间异质性^[14]。多种组学测序和单细胞测序技术的发展能更加深入了解HCC肿瘤的异质性^[15]。Losic等^[16]采集了来自14名HCC患者的71个区域样本，对其进行了DNA测序、RNA-seq和TCR-seq来研究HCC瘤内异质性。他们发现，在同一肿瘤内部的抗原表达，免疫浸润水平均有显著差异，其中区域性克隆免疫反应对HCC瘤内异质性的形成起着重要作用。为了以更高的分辨率了解这些数据，该团队对2名HCC患者中不同区域内的肿瘤进行单细胞测序，其结果进一步表明了即使在同一结节内距离较远的两个肿瘤组织，其激活的转录途径也存在着显著差异。

实际上，肿瘤异质性被认为是多种药物在癌症临床试验中失败以及对现有药物出现耐药性的主要原因。Gao等^[17]对10名接受根治性切除术的HCC患者的55个区域进行采样和低传代培养，并对其进行了DNA测序、拷贝数分析和高通量药物筛选。发现其中只有4个样本的亚区检测到*FGF19*、*DDR2*、*PDGFRA*和*TOP1*等基因改变并对相应的靶向治疗药物敏感。而索拉非尼作为目前全球针对晚期HCC的一线治疗药物^[18]，尽管它可以略微延长晚期HCC患者生存率，但由于不同患者的肿瘤之间相关药物转运蛋白的表达、细胞内药物代谢、信号通路的激活或失活、细胞内和细胞间特性等具有一定差异，使其在大量的患者中观察到其药效性受到了不同程度的影响^[19]。因此，如何利用关键特征将不同的HCC患者分为相对同质的亚型，具有重要临床意义。下一节将具体描述结合多种组学数据对HCC分层的策略及方法。

2 多组学HCC分层策略

精准肿瘤学的主要目的之一便是识别癌症分子亚型，即将具有共同生物学特征或临床表型（如生存时间和药物敏感性）的患者群体进行分类，使不同患者可以根据其所属亚型不同选择更加适合其自身的治疗方案。聚类是多年来相关研究人员在癌症分层中的常用算法。至今已开发出了许多聚类方法，如层次聚类、一致性聚类、基于密度、分布或质心的方法、半监督或监督的方法等^[20]。本文根据输入聚类算法组学数量，将当前多组学HCC分层策略分为从单组学出发（from single-to-multi, S To M）和从多组学出发（from multi-to-multi, M To M）两大策略。

2.1 从单组学出发（S To M）策略

S To M策略即利用单组学数据的不同特征对HCC进行分层后，结合多种组学数据寻找不同HCC亚型之间的差异分子，并验证其差异的真实性和肿瘤生物学现象的关联（如肿瘤的发生、转移、预后、免疫、代谢、信号通路等）（图1）。特征选择是S To M策略的核心，多年来针对分层特征的选择可大致分为基于数据分布的方法、基于生物特征的方法和基于多组学的方法3种。表1总结了近年来S To M策略下HCC分层研究的方法和结果。

2.1.1 基于数据分布的方法

基于数据分布的方法主要是根据数据分布特点或其在不同临床属性间的差异对特征进行筛选，其常见方法有方差筛选、变异系数筛选、中位偏差筛选、Cox模^[21]等。Ao等^[22]通过单因素Cox回归模型鉴定出192组HCC预后相关基因对后，选择了其中C指数最高的20组基因对，并根据其对患者的风险评分将患者分为高低风险组。综合多组学研究发现，两个亚型在基因突变和拷贝数变异上具有显著差异。其中，高风险组转录特征主要为细胞增殖和肿瘤微环境等相关途径的激活，而低风险患者的转录特征在于各种代谢途径的激活，并且两个亚型之间的转录差异与其DNA甲基化差异显著相关。Jiang等^[23]则选取变异系数前25%的蛋白质组学数据进行非负矩阵分解（nonnegative matrix factorization, NMF）和一致性聚类将110名早期HCC患者分为3个子类（SI、SII和SIII）。通过多种组学（蛋白质组、转录组和突变谱）分析发现，SI和SII具强增殖特征，并且*CTNNB1*的突变频率

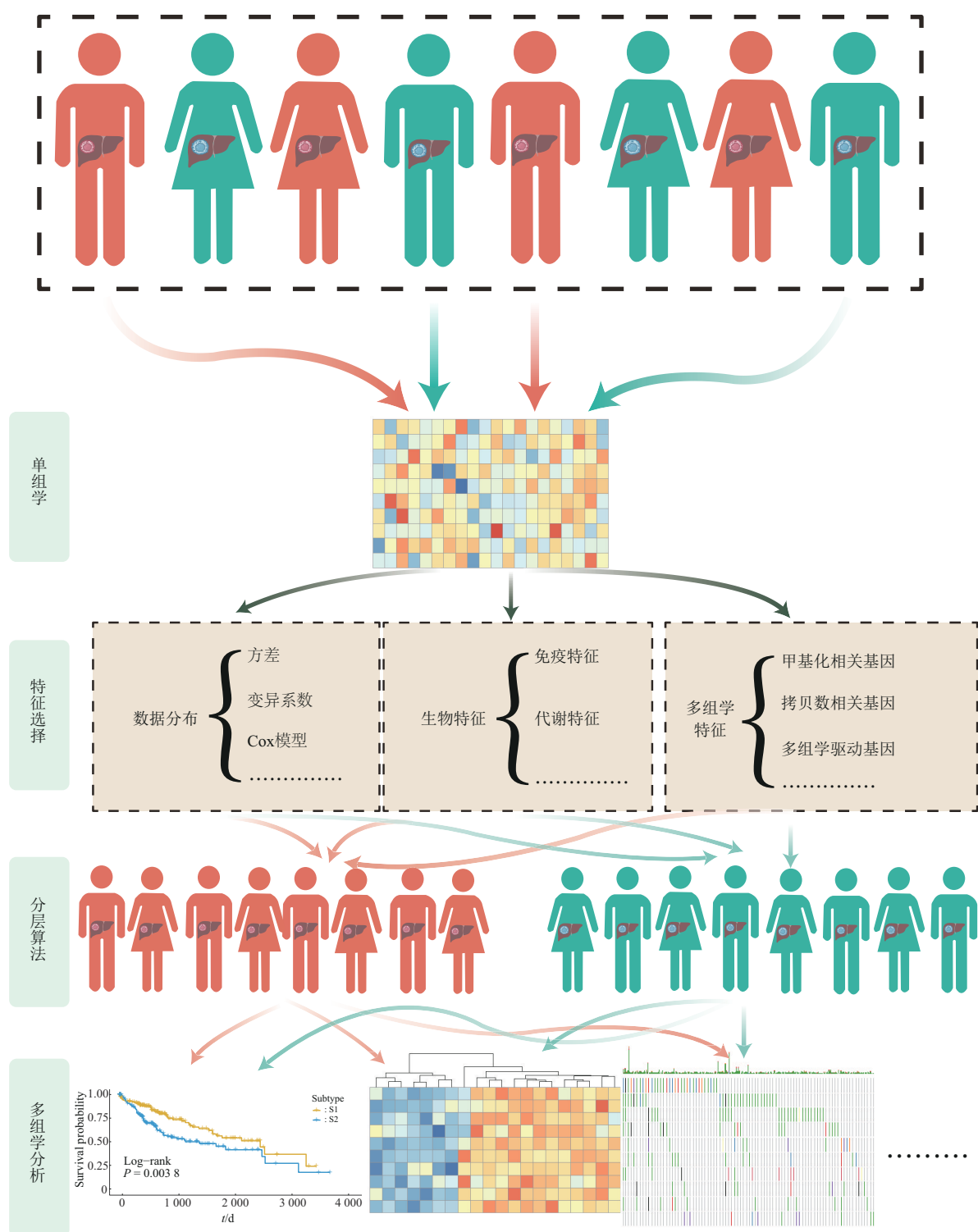


Fig. 1 Hepatocellular carcinoma stratification using S To M strategy
图1 从单一组学出发的肝细胞癌分层策略

和WNT通路的激活程度要高于SIII, 相比之下SIII则具有强侵袭性, 其富集了很多肿瘤促进通路, 并且存在抑制性免疫亚群。

2.1.2 基于生物特征的方法
随着对癌症研究的不断深入, 人们发现癌细胞可以激活不同的免疫检查点通路, 以达到抑制机体

免疫, 促进其生长发展的目的。针对PD-1/PD-L1、CTLA-4等免疫检查点的抑制剂可以重新激活抗癌免疫反应, 并在某些癌症的临床治疗上展现出良好的治疗效果^[24]。有研究表明, 在慢性肝炎的环境下, 调节性T细胞、髓源性抑制细胞的募集以及免疫检查点CTLA-4和PD-1的上调, 通常会促进HCC生长与发展^[25]。因此, 免疫检查点抑制剂似乎是有希望的HCC治疗策略。此外, 肝脏作为重要代谢器官, 其可以通过不同代谢物的生理调节来控制全身能量代谢, 其代谢异常也往往会引起如非酒精性脂肪肝和2型糖尿病等疾病^[26]。实际上, 一些代谢物的指标如葡萄糖和醋酸盐利用率、血清中乳酸含量已被认为是HCC有效的分层或预后指标^[27]。因此, 目前在HCC分层策略中, 针对生物学特征的提取主要为免疫特征和代谢特征。Zhang等^[28]对HCC免疫细胞质谱流式数据进行层次聚类, 将HCC分为3个亚型。其中亚型1为免疫活性亚型, 具有相对正常的T细胞浸润水平, 但B细胞浸润水平较低。亚型2为免疫缺陷亚型, 其特征在于淋巴细胞的浸润减少、树突状细胞和自然杀伤细胞的增加、单核细胞IL-1 β 过表达。亚型3为免疫抑制亚型, 其Treg细胞、Breg细胞和M2极化巨噬细胞等免疫抑制细胞的浸润水平要明显高于其他两个亚型, 并且一些免疫抑制分子如PD-1、PD-L1、

Tim-3和CTLA-4在该亚型中过表达。Bidkhoriet al^[29]使用转录组数据构建了一个由1 972个代谢基因组成的HCC特异性代谢网络。之后, 作者使用一种网络平滑的算法(network-based stratification, NBS)来传播每个基因的表达对其在网络相邻基因的影响^[30], 并将HCC分为3个子类。多种组学整合分析发现3种不同亚型的代谢和信号通路在基因组、转录组和蛋白质组水平上均具有显著差异, 并且3种亚型倾向依赖不同的同工酶来催化同一生化反应。

2.1.3 基于多组学的方法

除了上述方法之外, 还有一些分层特征本身是由多种组学确定的也被列入S To M策略中。例如Yang等^[31]对849名肝癌患者使用DriverNet算法^[32](一种基于突变谱和转录谱的驱动基因发现算法), 鉴定出34个HCC驱动基因。在后续的分析中作者将这34个驱动基因的突变谱输入NBS算法中^[30], 将HCC分为3个具有显著生存差异的子类(NBS1、NBS2和NBS3)。其中NBS1和NBS2富集了如TP53、AXIN1、RBI等抑癌基因, NBS3则具高CTNNB1突变频率和低抑癌基因突变频率, 并且该分类模型与先前报道的基于转录组的分类模型显著相关。

Table 1 The summary of methods and results for stratification of hepatocellular carcinoma by using the S To M strategy

表1 S To M策略肝细胞癌分层方法与结果小结

作者	年份	分层组学	特征选择	分层算法	最佳类数	分析组学	参考文献
Boyaulet al	2007	RNA	离差平方和和1-Pearson相关系数在样本中变化最大的基因	层次聚类	6	RNA、突变、甲基化、CNV	[33]
Miao等	2014	病理特征	根据患者临床病理评估和术后结果进行分类	人工划分	2	RNA、突变、CNV	[34]
Ao等	2016	RNA	单变量Cox模型中对预后具有显著影响的基因对(C指数TOP20)	人工划分高低风险组	2	RNA、突变、甲基化、CNV	[22]
Chaisaingmongkol等	2017	RNA	肿瘤较正常样本的差异基因	一致性聚类	3	RNA、突变、CNV、代谢组	[35]
Bidkhoriet al	2018	RNA	构建HCC特异性基因组规模代谢网络	网络传播+NMF一致性聚类	3	RNA、突变、甲基化	[29]
Li等	2019	RNA	778个免疫相关基因	一致性聚类	5	RNA、突变	[36]
Benfeitas等	2019	RNA	共表达分析鉴定了两个氧化还原簇G6PD和ALDH2	一致性聚类	2	RNA、代谢、突变	[37]
Jiang等	2019	蛋白组	变异系数TOP25%的蛋白质表达	NMF一致性聚类	3	RNA、突变、蛋白组	[23]
Gao等	2019	蛋白组	肿瘤相较正常样本的差异蛋白表达	一致性聚类	3	RNA、突变、CNV、蛋白质组	[38]

续表1							
作者	年份	分层组学	特征选择	分层算法	最佳类数	分析组学	参考文献
Shimada等	2019	RNA	肿瘤较正常样本差异表达基因, 且四分位距>1.5	一致性聚类+有无CTNNB1突变情况	3	RNA、突变、甲基化、CNV	[39]
Zhang等	2019	质谱流式细胞术	使用质谱流式细胞术获得HCC免疫细胞组成	层次聚类	3	RNA、代谢	[28]
Ouyang等	2020	RNA	联合Fisher比率, Spearman相关系数和决策树模型获得34个具有良好分类能力的基因	贝叶斯聚类	3	RNA、突变	[40]
Liu等	2020	RNA	利用表达数据计算肿瘤微环境细胞丰度	K-means	3	RNA、突变、甲基化、CNV、miRNA	[41]
Yang等	2020	RNA	具有高中位偏差和预后相关的代谢相关基因	NMF	3	RNA、突变、CNV	[42]
Deng等	2021	RNA	288个糖酵解相关基因	层次聚类	2	RNA、突变、甲基化、CNV	[43]
Yang等	2021	突变	基于RNA和突变数据鉴定驱动突变基因	网络传播+NMF一致性聚类	3	None	[31]
Xu等	2021	突变	由于体细胞突变数据计算的肿瘤突变负荷	人工划分高低风险组	2	RNA、突变	[44]
Wang等	2021	RNA	LASSO-Cox回归确定了5种与PD-L1和INF γ 强相关的免疫细胞	层次聚类	2	RNA、CNV、突变、甲基化、miRNA	[45]
Wei等	2021	RNA	使用RNA和突变数据鉴定癌症基因并选择排名TOP7和表达值标准偏差大于5的基因作为分层基因	K-means	2	—	[46]

RNA表示RNA-seq或微阵列数据; 突变: 体细胞突变数据; CNV: 拷贝数变异; 甲基化: DNA甲基化数据; 蛋白质组: 蛋白质组学和磷酸蛋白质组学数据; —: 论文未做相关分析。

2.2 从多组学出发 (M To M) 策略

M To M策略则希望通过对多种组学数据 (如基因组、转录组、表观基因组与肿瘤生物学的关联) 进行多层面的综合分析 (即从系统生物学的概念出发) 全景式地展示不同亚型内各组学之间的差异与联系。数据降维是M To M策略的核心步骤, 即将多种组学数据集中放入一个低维的堆叠矩阵中, 然后再对该数据矩阵实施聚类算法以获得具有多种组学特征的分层结果 (图2)。根据使用的算法不同, 常见的M To M策略分层算法可以分为3类, 基于相似性的方法、基于集成的方法和深度学习。表2总结了近年来M To M策略下HCC分层研究的方法和结果。

2.2.1 基于相似性的方法

基于相似性的方法主要思想是将不同的组学特征转化为患者之间的相似程度, 并输出一个综合的“患者-患者相似矩阵”, 最终根据该矩阵得到患者分层结果。相似网络融合 (similarity network

fusion, SNF) 是这一类方法的代表性算法^[47], 该算法首先将每个输入的组学数据转化为“患者-患者相似网络”。在此网络中节点表示每个患者, 连接边上的权值则表示患者之间相似度的大小, 之后通过融合迭代公式将多个相似网络逐渐融合。当达到公式收敛条件时得到最终的“融合相似网络”, 并在此基础上完成患者分层工作。经过多年发展, SNF算法及其改进算法常常被用于HCC多组学分层研究中。Ruan等^[48]提出了一种加强信号注释的相似网络融合 (association-signal-annotation boosted SNF, ab-SNF) 模型, 与原始SNF算法相比, ab-SNF将不同组学数据特征和感兴趣结果之间的关联信号注释作为权值, 加入到构建患者之间相似网络中, 以减少噪声的影响并提高聚类性能。在大型患者队列中, 部分患者某项组学数据的丢失是一个常见的问题。为解决这一问题, Xu等^[49]提出了多重相似网络嵌入 (multiple similarity network embedding, MSNE) 模型, MSNE算法的

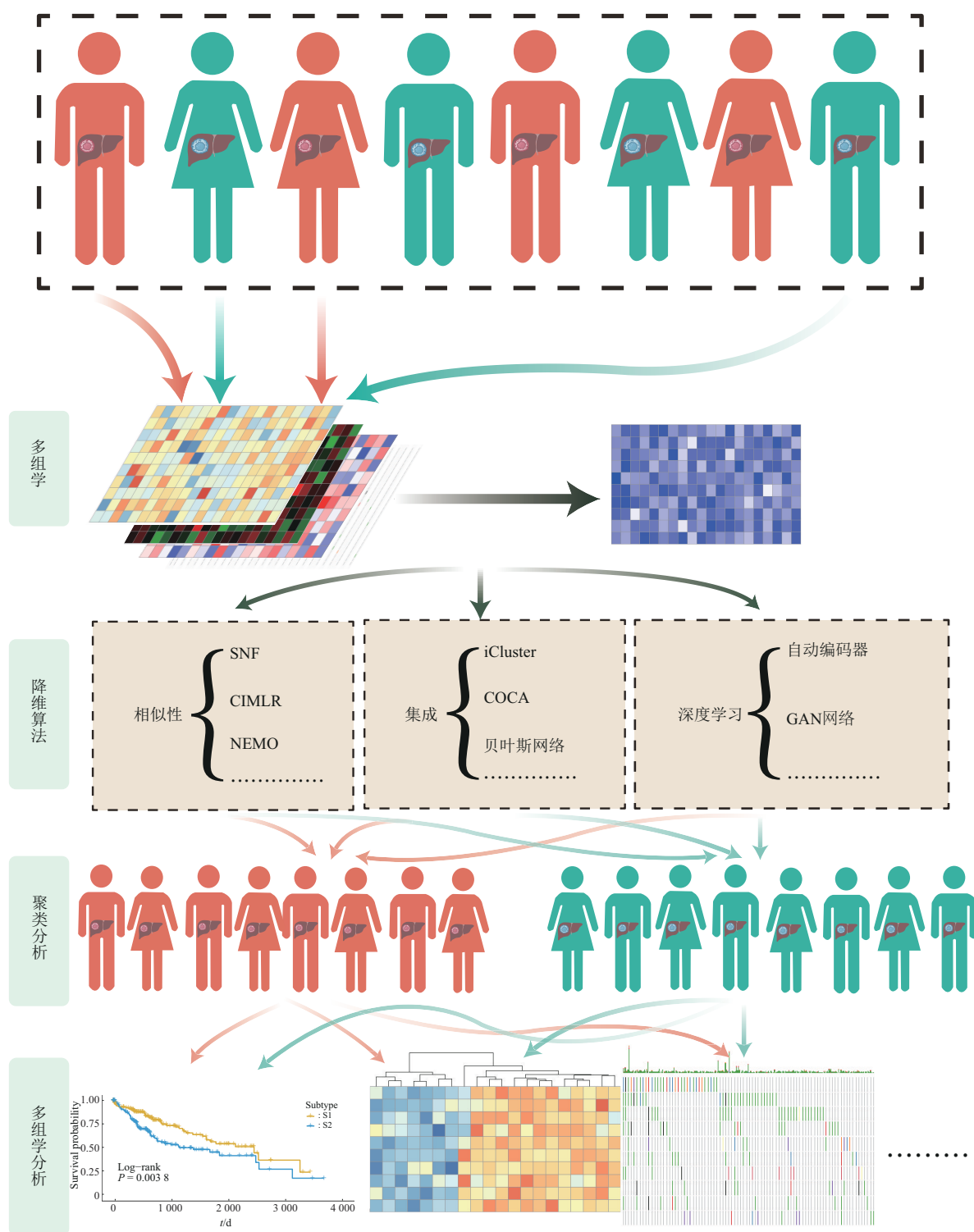


Fig. 2 Hepatocellular carcinoma stratification using M To M strategy

图2 从多组学出发的肝细胞癌分策略

原理是在构建完单个患者相似网络后，采用随机游走的方式从多个网络中获得一个综合相似网络，使一些组学数据丢失的样本，也可以被投影到低维的

相似网络中。并且相对于原始SNF算法，MSNE算法得到的结果具有更丰富的临床参数和更显著的生存差异。

除了SNF算法外, 其他基于相似性的方法也用于HCC分层研究中。例如Ramazzotti等^[50]开发了一种多核学习亚型识别算法 (multikernel learning, CIMLR), 该算法可以根据每种组学在不同癌症中的重要程度为其分配权重, 并且利用每个组学的多个高斯核构建“患者-患者相似矩阵”。在实际应用中, CIMLR将来自TCGA的359例HCC样本分为了8个亚型, 其结果具有显著的疾病特异性和生存差异。基于邻域的多组学亚型识别算法 (neighborhood based multi-omics clustering, NEMO)^[51]模型认为每个样本的局部邻域特征, 可以更好地捕捉患者在每个组学中的相似模式, 其大致可分为3个步骤: a. 为每个组学构建患者间相似矩阵; b. 将来自不同组学的相似矩阵整合到一个矩阵中; c. 对该矩阵进行聚类分析。NEMO在应用中不需要迭代优化, 具有比SNF类算法更快的执行速度, 且无需对丢失数据的样本进行插补或删除。

2.2.2 基于集成的方法

基于贝叶斯框架、主成分分析^[52]、矩阵分解^[53]等对数据或模型集成的方法也常被用于癌症亚型识别研究当中。例如基于联合潜变量模型的iCluster, 该模型假设肿瘤亚型为未观察到的潜在变量, 并且该变量会形成一组低维的空间坐标, 可以捕捉不同组学之间的相关性并用于肿瘤样本的聚类中^[54]。值得一提的是, 最初的iCluster仅能输入连续变量 (如表达数据、DNA甲基化数据), 经过不断的改进其最新版本iClusterBayes采用全贝叶斯潜变量模型, 不但允许输入二值变量 (如突变数据)、分类变量 (如基因拷贝数状态) 和连续变量, 还极大地减少了算法运行时间^[55]。在实际应

用中, 一些研究表明iCluster算法可以将HCC分为稳定且具有显著临床差异的3个亚型^[56-57]。

除了输入特定的组学数据以外, 聚类分配 (cluster-of-cluster-assignments, COCA) 算法可以允许输入单组学聚类结果, 并从中分配得出二级聚类结果^[58]。Yang等^[59]探索了来自不同组学的HCC驱动因素, 并使用COCA算法对其进行综合聚类, 得到了4个稳定的HCC亚型 (C1、C2、C3和C4)。其中C1肿瘤主要富集了DNA修复和病毒致癌通路的异常, C2肿瘤的特征主要在于NF-κB通路和NBEA的突变, C3和C4肿瘤中特异性表达的基因则主要与免疫应答和T细胞调节相关。

2.2.3 深度学习

最近, 随着人工智能领域的不断发展, 深度学习作为这一领域的热点, 在医学影像、信号及组学数据的处理中得到了广泛应用^[60]。自动编码器作为常见的人工神经网络框架, 常被用于对一组数据的特征学习和降维工作中。并且在医学组学数据中, 该算法被证明可以有效提取与临床和分子有关的特征^[61]。Chaudhary等^[62]对来自TCGA的360例HCC样本的RNA-seq、miRNA-Seq和DNA甲基化数据应用自动编码器降维后, 使用单因素Cox模型提取了37个与生存显著相关的临床特征, 并对这些特征使用K-means聚类得到了两个具有显著生存差异的亚型。其中侵袭性亚型S1的差异基因主要集中在癌症相关通路、Wnt信号通路、PI3K-Akt信号通路等, 而中度侵袭性亚型S2则主要为代谢相关通路的激活, 如药物代谢、氨基酸和脂肪酸代谢等。Wang等^[63]则将自动编码器与SNF算法相结合, 同样地得到了两个具有显著生存差异的HCC亚类。

Table 2 The summary of methods and results for stratification of hepatocellular carcinoma by using M To M strategy

表2 M To M策略肝细胞癌分层方法与结果小结

作者	年份	分层组学	聚类方法	计算环境	类别	最佳类数	参考文献
Ramazzotti等	2018	RNA、突变、甲基化、CNV	多核学习模型 (CIMLR)	R/Matlab	基于相似性	8	[50]
Rappoport等	2019	RNA、miRNA、甲基化	基于邻域模型 (NEMO)	R	基于相似性	5	[51]
Ruan等	2019	RNA、突变、甲基化	加强信号注释的相似网络融合 (ab-SNF)	R	基于相似性	5	[48]
Luo等	2019	RNA、miRNA、突变、甲基化、CNV	随机森林聚类	R	基于相似性	2	[64]
Liu等	2020	RNA、miRNA、甲基化	基于生存的图聚类模型	Matlab	基于相似性	3	[65]
Rappoport等	2020	RNA、miRNA、甲基化	非穷举模式的聚类模型 (MONET)	R/Python	基于相似性	5	[66]

续表2

作者	年份	分层组学	聚类方法	计算环境	类别	最佳类数	参考文献
Xu等	2021	RNA、甲基化、CNV	多重相似网络嵌入 (MSNE)	Python	基于相似性	文章未给出最佳类别	[49]
Feng等	2021	RNA、miRNA、甲基化	相似核学习模型	作者并未进行说明	基于相似性	7	[67]
Wang等	2021	RNA、miRNA、甲基化	去噪网络正则化模型 (DeFusion)	R/Python/Matlab	基于相似性	6	[68]
TCGA	2017	RNA、miRNA、甲基化、CNV、RPPA	基于联合潜变量模型 (iCluster)	R	基于集成	3	[56]
Woo等	2017	RNA、甲基化、CNV	基于联合潜变量模型 (iCluster)	R	基于集成	3	[57]
Yang等	2020	RNA、miRNA、甲基化、CNV	聚类分配算法 (COCA)	R	基于集成	4	[59]
Suter等	2021	RNA、miRNA、甲基化、CNV、突变	贝叶斯网络模型	R	基于集成	3	[69]
Chaudhary等	2018	RNA、miRNA、甲基化	自动编码器	Python	深度学习	2	[62]
Wang等	2020	RNA、miRNA、甲基化	自动编码器+基于径向基核的相似网络融合	Python	深度学习	2	[63]
Poirion等	2021	RNA、miRNA、甲基化	自动编码器	Python	深度学习	2	[70]

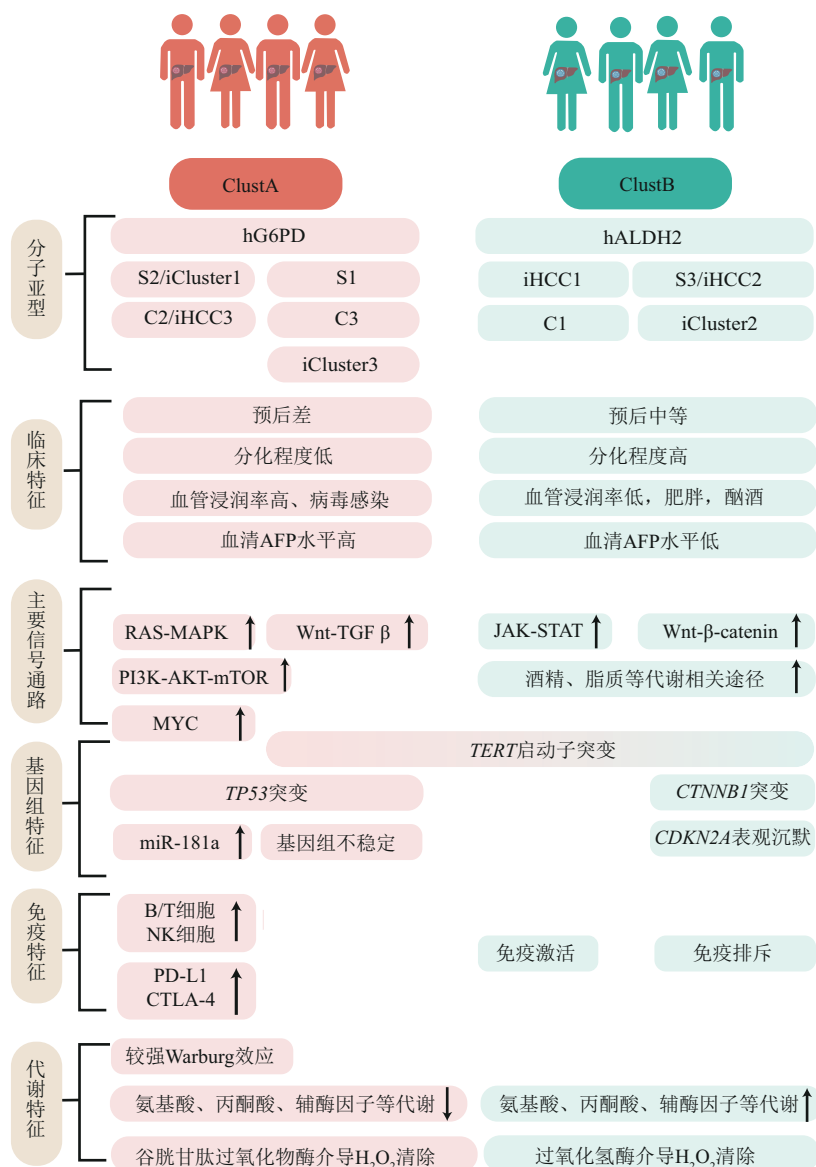
RNA和miRNA表示RNA-seq/miRNA-seq或微阵列数据; 突变: 体细胞突变数据; CNV: 拷贝数变异; 甲基化: DNA甲基化数据; RPPA: 反相蛋白微阵列数据。

3 多组学HCC亚型特点

值得注意的是, 一些HCC亚型在很多独立的亚型分类研究中反复出现, 这暗示着来源于不同方法的HCC亚型可能具有共同的特征。本小节综合了Hoshida等^[71]、Ally等^[56]、Yang等^[42]、Benfeitas等^[37]、Bidkhor等^[29]以及两篇相关综述^[7, 9], 将HCC大致分成两类: ClustA和ClustB (图3)。其中ClustA更具侵袭性, 在临床上组织分化程度低、血管浸润率和患者血清中甲胎蛋白(AFP)水平偏高。在基因组特征上ClustA表现出了更加频繁的TP53突变, 染色体高度不稳定性以及一些常见的致癌通路如PI3K-AKT-mTOR、RAS-MAPK、WNT等的高度激活。此外, 部分ClustA样本具有免疫耗竭特征, 具体以TGF- β 、PD-L1、CTLA-4等驱动的T细胞耗竭状态为主^[42]。

相反, ClustB具有相对ClustA而言较好的组织分化能力、较低的AFP水平、较好的预后, 并且该亚类多与过度肥胖和过度酗酒有关。因此, 可以

发现有关酒精、脂质等代谢通路在ClustB中高度激活。Wnt- β -catenin信号的持续激活在HCC中是一个频繁发生的驱动事件^[72]。值得一提的是, 两种亚型似乎会以不同的方式激活Wnt通路。其中ClustB主要以CTNNB1的突变激活Wnt通路, 这种类型的样本在一些研究中也证明与更好的预后相关^[73]。相反, 在侵袭性表型ClustA中则发现了TGF- β 的过度表达, 其主要通过调节胞内游离的 β -catenin来增强Wnt通路^[71]。此外, 目前普遍认为癌症细胞在高速代谢生长的同时, 会促进大量NADPH的生物合成以抵抗活性氧(ROS)对其自身的伤害^[74]。然而有趣的是, 这两种HCC表型被发现会以不同的方式清除胞内ROS。其中, 弱侵袭性ClustB主要显示由过氧化氢酶介导的ROS清除, 另一种亚型则主要以谷胱甘肽过氧化物酶依赖的方式清除ROS^[37]。这似乎也表明了针对HCC的抗氧化剂治疗中, 需要根据不同的亚型选择不同的药物。

**Fig. 3 Multi-omics characteristics of HCC subtypes****图3 HCC亚型的多组学特点**

Hoshida等^[71]分类: S1、S2和S3; TCGA^[56]分类: iCluster1、iCluster2和iCluster3; Yang等^[42]分类: C1、C2和C3; Benfeitas等^[37]分类: hG6PD和hALDH2; Bidkhor等^[29]分类: iHCC1、iHCC2和iHCC3; ↑: 激活/表达上升; ↓: 抑制/表达下降。

4 总结与讨论

HCC在分子和病理上的高度异质性极大地阻碍了其临床治疗效果。因此,将HCC患者分为相对同质的亚型,对其临床治疗效果和个性化治疗有着重要意义。随着高通量技术的不断发展,多种组学数据综合分析可以帮助研究人员更好地了解HCC发展背后的生物学机制,也为HCC分层研究打开了新的思路。

根据分层算法输入组学的数量,本文将目前HCC多组学分层方法分为S To M和M To M两大策略。其中S To M策略是利用单组学数据的不同特征对HCC进行分层后,结合多种组学数据寻找不同HCC亚型之间的差异分子,并验证其差异的真实性与肿瘤生物学现象的关联。特征选择是S To M策略的核心,其中,基于生物学特征的方法,主要聚焦在将不同的疾病表型与特定基因型的关联,有助于其后续的转化研究,是目前S To M分层策略

的主要手段（例如，在表1的19项研究中有9项研究选择HCC的生物学特征对其进行分层）。免疫特征和代谢特征是目前HCC分层研究中常用的生物学特征。然而，随着人们对癌症的认识不断更新，一些新的癌症标志（如细胞衰老、非突变表观遗传、多态微生物组等）也被认为对形成恶性肿瘤起到至关重要的作用^[75]。我们认为这些标志同样具有良好的分层潜力，可以用于HCC亚型分类研究中，以求通过不同的角度全面了解HCC异质性及其分子特点。此外，从近几年相关研究来看，连续型分层特征（如转录组、蛋白质组）在其中占据了绝大多数，而离散特征（如体细胞突变）则很少作为HCC的分层特征。这很大程度是因为无法通过离散特征计算出患者之间的欧几里得距离，因此不能满足当前分层算法（如层次聚类、K-means、一致性聚类等）的输入要求。然而Yang等^[31]和Xu等^[44]的研究让我们看到了通过对离散特征进行适当的去稀疏化处理后，其具有不亚于连续特征的分层能力。考虑到癌症组学的离散特征在其发展中的有着重要意义，并且在临床上其操作相对简单，在以后的研究中可以考虑在分层特征中加入患者的组学的离散特征，以获得更具全面分层结果。

M To M策略则是从系统生物学的概念出发，全景式地展示不同亚型内各组学之间的差异与联系。然而，与S to M相比，M To M需要用户掌握多种编程语言（表2）以及更强大的计算性能。并且，不同组学数据处理方法的成熟度参差不齐，也是当前多组学集成结果转化为临床解释的主要障碍^[76]。因此，加强样本处理和相关分析方法的标准化流程以及开发稳定健壮的多组学集成工具，对促进相关理论的发现和结果的可翻译性至关重要。此外，需要注意的是，一般认为合并更多的组学数据往往会得到更好的分层结果。而Duan等^[77]的研究驳斥了这个观点，他们的研究发现，在某些情况下集成更多的组学内容反而会对结果造成负面影响。因此，使用哪几种组学数据的组合可以有效地完成HCC分层任务，是以后使用M To M策略下相关算法前需要考虑的。其次，在大型患者队列中，数据丢失是十分常见的现象，单纯对丢失数据的样本进行删除可能会对其结果的统计学指标造成影响。因此，如何对丢失数据做出适当的处理也在对HCC多组学分层前需要考虑的。组学分析需要的数据量大，往往需要多个中心的合作。整合来自多个中心或研究的数据集可以获得更可靠的结果和

潜在的新发现^[78]。但在整合研究时，也需要注意由实验室、操作人员、操作平台等其他非生物因素的差异引起的批次效应可能会掩盖或降低未发现差异信号的强度^[79]。

最后，如Ally等^[56]、Hoshida等^[71] HCC亚型在很多相关的亚型研究中反复出现，这表明了使用不同方法得到的HCC亚型可能具有共同特征。然而，针对当前HCC独立亚型之间相似性的研究仍十分浅薄^[9]。因此，未来仍需进行更多的研究，以在其中总结出更具有代表性的亚类。

参 考 文 献

- [1] Sung H, Ferlay J, Siegel R L, *et al.* Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*, 2021, **71**(3): 209-249
- [2] Bertuccio P, Turati F, Carioli G, *et al.* Global trends and predictions in hepatocellular carcinoma mortality. *J Hepatol*, 2017, **67**(2): 302-309
- [3] Cao M, Li H, Sun D, *et al.* Cancer burden of major cancers in China: a need for sustainable actions. *Cancer Commun (Lond)*, 2020, **40**(5): 205-210
- [4] Balogh J, Victor Iii D, Asham E H, *et al.* Hepatocellular carcinoma: a review. *J Hepatocell Carcinoma*, 2016, **3**: 41-53
- [5] Mcglynn K A, Petrick J L, El-Serag H B. Epidemiology of hepatocellular carcinoma. *Hepatology*, 2021, **73**: 4-13
- [6] Craig A J, Von Felden J, Garcia-Lezana T, *et al.* Tumour evolution in hepatocellular carcinoma. *Nat Rev Gastroenterol Hepatol*, 2020, **17**(3): 139-152
- [7] Llovet J M, Kelley R K, Villanueva A, *et al.* Hepatocellular carcinoma. *Nat Rev Dis Primers*, 2021, **7**(1): 6
- [8] Schulze K, Imbeaud S, Letouzé E, *et al.* Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. *Nat Genet*, 2015, **47**(5): 505-511
- [9] Wu Y, Liu Z, Xu X. Molecular subtyping of hepatocellular carcinoma: a step toward precision medicine. *Cancer Commun (Lond)*, 2020, **40**(12): 681-693
- [10] 国家卫生健康委办公厅. 原发性肝癌诊疗指南(2022年版). 临床肝胆病杂志, 2022, **38**(2): 288-303
General Office of National Health Commission. Standard for diagnosis and treatment of primary liver cancer. *J Clin Hepatol*, 2022, **38**(2): 288-303
- [11] Tsilimigras D I, Bagante F, Sahara K, *et al.* Prognosis after resection of barcelona clinic liver cancer (BCLC) stage 0, A, and B hepatocellular carcinoma: a comprehensive assessment of the current BCLC classification. *Ann Surg Oncol*, 2019, **26**(11): 3693-3700
- [12] Chakraborty S, Hosen M, Ahmed M, *et al.* Onco-multi-omics approach: a new frontier in cancer research. *Biomed Res Int*, 2018, **2018**: 9836256

- [13] Nicora G, Vitali F, Dagliati A, *et al.* Integrated multi-omics analyses in oncology: a review of machine learning methods and tools. *Front Oncol*, 2020, **10**: 1030
- [14] Lu L C, Hsu C H, Hsu C, *et al.* Tumor heterogeneity in hepatocellular carcinoma: facing the challenges. *Liver Cancer*, 2016, **5**(2): 128-138
- [15] Dhanasekaran R. Deciphering tumor heterogeneity in hepatocellular carcinoma (HCC) —multi-omic and singulomic approaches. *Semin Liver Dis*, 2021, **41**(1): 9-18
- [16] Losic B, Craig A J, Villacorta-Martin C, *et al.* Intratumoral heterogeneity and clonal evolution in liver cancer. *Nat Commun*, 2020, **11**(1): 291
- [17] Gao Q, Wang Z C, Duan M, *et al.* Cell culture system for analysis of genetic heterogeneity within hepatocellular carcinomas and response to pharmacologic agents. *Gastroenterology*, 2017, **152**(1): 232-242.e234
- [18] Llovet J M, Ricci S, Mazzaferro V, *et al.* Sorafenib in advanced hepatocellular carcinoma. *N Engl J Med*, 2008, **359**(4): 378-390
- [19] Cabral L K D, Tiribelli C, Sukowati C H. Sorafenib resistance in hepatocellular carcinoma: the relevance of genetic heterogeneity. *Cancers (Basel)*, 2020, **12**(6): 1576
- [20] Brière G, Darbo É, Thébault P, *et al.* Consensus clustering applied to multi-omics disease subtyping. *BMC Bioinformatics*, 2021, **22**(1): 1-29
- [21] Cox D R. Regression models and life-tables. *J R Stat Soc Series B Stat Methodol*, 1972, **34**(2): 187-202
- [22] Ao L, Song X, Li X, *et al.* An individualized prognostic signature and multi-omics distinction for early stage hepatocellular carcinoma patients with surgical resection. *Oncotarget*, 2016, **7**(17): 24097
- [23] Jiang Y, Sun A, Zhao Y, *et al.* Proteomics identifies new therapeutic targets of early-stage hepatocellular carcinoma. *Nature*, 2019, **567**(7747): 257-261
- [24] Darvin P, Toor S M, Sasidharan Nair V, *et al.* Immune checkpoint inhibitors: recent progress and potential biomarkers. *Exp Mol Med*, 2018, **50**(12): 1-11
- [25] Donisi C, Puzzone M, Ziranu P, *et al.* Immune checkpoint inhibitors in the treatment of HCC. *Front Oncol*, 2021, **10**: 601240
- [26] Rui L. Energy metabolism in the liver. *Compr Physiol*, 2014, **4**(1): 177
- [27] De Matteis S, Ragusa A, Marisi G, *et al.* Aberrant metabolism in hepatocellular carcinoma provides diagnostic and therapeutic opportunities. *Oxid Med Cell Longev*, 2018, **2018**: 7512159
- [28] Zhang Q, Lou Y, Yang J, *et al.* Integrated multiomic analysis reveals comprehensive tumour heterogeneity and novel immunophenotypic classification in hepatocellular carcinomas. *Gut*, 2019, **68**(11): 2019-2031
- [29] Bidkhori G, Benfeitas R, Klevisig M, *et al.* Metabolic network-based stratification of hepatocellular carcinoma reveals three distinct tumor subtypes. *Proc Natl Acad Sci USA*, 2018, **115**(50): E11874-E11883
- [30] Hofree M, Shen J P, Carter H, *et al.* Network-based stratification of tumor mutations. *Nat Methods*, 2013, **10**(11): 1108-1115
- [31] Yang C, Guo Y, Qian R, *et al.* Mapping the landscape of synthetic lethal interactions in liver cancer. *Theranostics*, 2021, **11**(18): 9038
- [32] Bashashati A, Haffari G, Ding J, *et al.* DriverNet: uncovering the impact of somatic driver mutations on transcriptional networks in cancer. *Genome Biol*, 2012, **13**(12): R124
- [33] Boyault S, Rickman D S, De Reyniès A, *et al.* Transcriptome classification of HCC is related to gene alterations and to new therapeutic targets. *Hepatology*, 2007, **45**(1): 42-52
- [34] Miao R, Luo H, Zhou H, *et al.* Identification of prognostic biomarkers in hepatitis B virus-related hepatocellular carcinoma and stratification by integrative multi-omics analysis. *J Hepatol*, 2014, **61**(4): 840-849
- [35] Chaisaingmongkol J, Budhu A, Dang H, *et al.* Common molecular subtypes among Asian hepatocellular carcinoma and cholangiocarcinoma. *Cancer Cell*, 2017, **32**(1): 57-70.e53
- [36] Li W, Wang H, Ma Z, *et al.* Multi-omics analysis of microenvironment characteristics and immune escape mechanisms of hepatocellular carcinoma. *Front Oncol*, 2019, **9**: 1019
- [37] Benfeitas R, Bidkhori G, Mukhopadhyay B, *et al.* Characterization of heterogeneous redox responses in hepatocellular carcinoma patients using network analysis. *EBioMedicine*, 2019, **40**: 471-487
- [38] Gao Q, Zhu H, Dong L, *et al.* Integrated proteogenomic characterization of HBV-related hepatocellular carcinoma. *Cell*, 2019, **179**(2): 561-577.e522
- [39] Shimada S, Mogushi K, Akiyama Y, *et al.* Comprehensive molecular and immunological characterization of hepatocellular carcinoma. *EBioMedicine*, 2019, **40**: 457-470
- [40] Ouyang X, Fan Q, Ling G, *et al.* Identification of diagnostic biomarkers and subtypes of liver hepatocellular carcinoma by multi-omics data analysis. *Genes*, 2020, **11**(9): 1051
- [41] Liu F, Qin L, Liao Z, *et al.* Microenvironment characterization and multi-omics signatures related to prognosis and immunotherapy response of hepatocellular carcinoma. *Exp Hematol Oncol*, 2020, **9**: 10
- [42] Yang C, Huang X, Liu Z, *et al.* Metabolism-associated molecular classification of hepatocellular carcinoma. *Mol Oncol*, 2020, **14**(4): 896-913
- [43] Deng T, Ye Q, Jin C, *et al.* Identification and validation of a glycolysis-associated multiomics prognostic model for hepatocellular carcinoma. *Aging (Albany NY)*, 2021, **13**(5): 7481
- [44] Xu Q, Xu H, Deng R, *et al.* Multi-omics analysis reveals prognostic value of tumor mutation burden in hepatocellular carcinoma. *Cancer Cell Int*, 2021, **21**(1): 342
- [45] Wang Z, Zhang S. Multi-omic analyses of hepatocellular carcinoma to determine immunological characteristics and key nodes in gene-expression network. *Biosci Rep*, 2021, **41**(7): BSR20211241
- [46] Wei T, Fa B, Luo C, *et al.* An efficient and easy-to-use network-based Integrative method of multi-omics data for cancer genes discovery. *Front Genet*, 2021, **11**: 613033

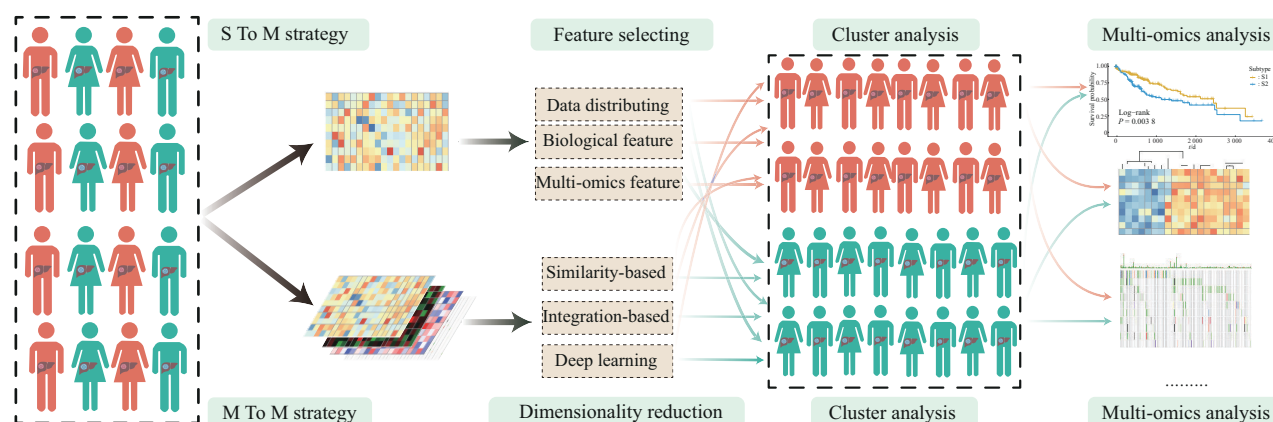
- [47] Wang B, Mezlini A M, Demir F, *et al.* Similarity network fusion for aggregating data types on a genomic scale. *Nat Methods*, 2014, **11**(3): 333-337
- [48] Ruan P, Wang Y, Shen R, *et al.* Using association signal annotations to boost similarity network fusion. *Bioinformatics*, 2019, **35**(19): 3718-3726
- [49] Xu H, Gao L, Huang M, *et al.* A network embedding based method for partial multi-omics integration in cancer subtyping. *Methods*, 2021, **192**: 67-76
- [50] Ramazzotti D, Lal A, Wang B, *et al.* Multi-omic tumor data reveal diversity of molecular mechanisms that correlate with survival. *Nat Commun*, 2018, **9**(1): 445
- [51] Rappoport N, Shamir R. NEMO: cancer subtyping by integration of partial multi-omic data. *Bioinformatics*, 2019, **35**(18): 3348-3356
- [52] Meng C, Helm D, Frejno M, *et al.* moCluster: identifying joint patterns across multiple omics data sets. *J Proteome Res*, 2016, **15**(3): 755-765
- [53] Liu J, Wang C, Gao J, *et al.* Multi-view clustering via joint nonnegative matrix factorization//Ghosh J, Obradovic Z, Dy J, *et al.* Proceedings of the 2013 SIAM International Conference on Data Mining (SDM). Philadelphia: SIAM, 2013: 252-260
- [54] Shen R, Olshen A B, Ladanyi M. Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis. *Bioinformatics*, 2009, **25**(22): 2906-2912
- [55] Mo Q, Shen R, Guo C, *et al.* A fully bayesian latent variable model for integrative clustering analysis of multi-type omics data. *Biostatistics*, 2018, **19**(1): 71-86
- [56] Ally A, Balasundaram M, Carlsen R, *et al.* Comprehensive and integrative genomic characterization of hepatocellular carcinoma. *Cell*, 2017, **169**(7): 1327-1341.e1323
- [57] Woo H G, Choi J-H, Yoon S, *et al.* Integrative analysis of genomic and epigenomic regulation of the transcriptome in liver cancer. *Nat Commun*, 2017, **8**(1): 839
- [58] Hoadley K A, Yau C, Wolf D M, *et al.* Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell*, 2014, **158**(4): 929-944
- [59] Yang L, Zhang Z, Sun Y, *et al.* Integrative analysis reveals novel driver genes and molecular subclasses of hepatocellular carcinoma. *Aging (Albany NY)*, 2020, **12**(23): 23849
- [60] Min S, Lee B, Yoon S. Deep learning in bioinformatics. *Brief Bioinform*, 2017, **18**(5): 851-869
- [61] Tan J, Ung M, Cheng C, *et al.* Unsupervised feature construction and knowledge extraction from genome-wide assays of breast cancer with denoising autoencoders//Altman B R, Dunker K A, Hunter L, *et al.* Pacific Symposium on Biocomputing 2015. Hawaii: World Scientific, 2014: 132-143
- [62] Chaudhary K, Poirion O B, Lu L, *et al.* Deep learning-based multi-omics integration robustly predicts survival in liver cancer. *Clin Cancer Res*, 2018, **24**(6): 1248-1259
- [63] Wang Z, Yan R, Liu J, *et al.* An integration framework for liver cancer subtype classification and survival prediction based on multi-omics data//Huang S D, Premaratne P. International Conference on Intelligent Computing 16th International Conference, ICIC 2020, Bari, Italy, October 2-5, 2020, Proceedings, Part III. Bari: Springer, 2020: 247-257
- [64] Luo Z, Wang W, Li F, *et al.* Pan-cancer analysis identifies telomerase-associated signatures and cancer subtypes. *Mol Cancer*, 2019, **18**(1): 106
- [65] Liu C, Wenming C, Wu S, *et al.* Supervised graph clustering for cancer subtyping based on survival analysis and integration of multi-omic tumor data. *IEEE/ACM Trans Comput Biol Bioinform*, 2020, **19**(2): 1193-1202
- [66] Rappoport N, Safra R, Shamir R. MONET: multi-omic module discovery by omic selection. *PLoS Comput Biol*, 2020, **16**(9): e1008182
- [67] Feng J, Jiang L, Li S, *et al.* Multi-omics data fusion via a joint kernel learning model for cancer subtype discovery and essential gene identification. *Front Genet*, 2021, **12**: 647141
- [68] Wang W, Zhang X, Dai D Q. DeFusion: a denoised network regularization framework for multi-omics integration. *Brief Bioinform*, 2021, **22**(5): bbab057
- [69] Suter P, Kuipers J, Beerenwinkel N. Discovering gene regulatory networks of multiple phenotypic groups using dynamic Bayesian networks. *Brief Bioinform*, 2022, **23**(4): bbac219
- [70] Poirion O B, Jing Z, Chaudhary K, *et al.* DeepProg: an ensemble of deep-learning and machine-learning models for prognosis prediction using multi-omics data. *Genome Med*, 2021, **13**(1): 112
- [71] Hoshida Y, Nijman S, Kobayashi M, *et al.* Integrative transcriptome analysis reveals common molecular subclasses of human hepatocellular carcinoma. *Cancer Res*, 2009, **69**(18): 7385-7392
- [72] Takigawa Y, Brown A. Wnt signaling in liver cancer. *Curr Drug Targets*, 2008, **9**(11): 1013-1024
- [73] Khalaf A M, Fuentes D, Morshid A I, *et al.* Role of Wnt/ β -catenin signaling in hepatocellular carcinoma, pathogenesis, and clinical significance. *J Hepatocell Carcinoma*, 2018, **5**: 61
- [74] Patra K C, Hay N. The pentose phosphate pathway and cancer. *Trends Biochem Sci*, 2014, **39**(8): 347-354
- [75] Hanahan D. Hallmarks of cancer: new dimensions. *Cancer Discov*, 2022, **12**(1): 31-46
- [76] Menyhárt O, Györfy B. Multi-omics approaches in cancer research with applications in tumor subtyping, prognosis, and diagnosis. *Comput Struct Biotechnol J*, 2021, **19**: 949-960
- [77] Duan R, Gao L, Gao Y, *et al.* Evaluation and comparison of multi-omics data integration methods for cancer subtyping. *PLoS Comput Biol*, 2021, **17**(8): e1009224
- [78] Jiang P, Sinha S, Aldape K, *et al.* Big data in basic and translational cancer research. *Nat Rev Cancer*, 2022, **22**(11): 625-635
- [79] Leek J T, Scharpf R B, Bravo H C, *et al.* Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet*, 2010, **11**(10): 733-739

The Strategies and Progression in The Stratification of Hepatocellular Carcinoma Using Multi-omics Data *

WANG Meng, LI Xiao-Qin **, GAO Bin

(Faculty of Environment and Life of Beijing University of Technology, Beijing 100124, China)

Graphical abstract



Abstract Hepatocellular carcinoma (HCC) is the most common and lethal liver malignancy. The treatment of this disease has been hampered by its heterogeneity, which severely limits progress in its personalized therapy. Therefore, it is necessary to divide highly heterogeneous HCC into molecular subtypes with similar characteristics for its clinical treatment. With the development of high-throughput technologies, integrative multi-omics data can deepen our comprehension of the biological mechanisms behind HCC pathogenesis. And it can also open new ideas for HCC stratification studies. Cluster analysis has been the main algorithm of cancer subtypes research for many years. Based on the number of input clustering algorithm omics, we summarize the current multi-omics HCC stratification methods into two major strategies: “from single-to-multi (S To M)” and “from multi-to-multi (M To M)”. Among them, the S To M strategy is to stratify HCC using different features of single omics and then combine multi-omics data to find the differential molecules among different HCC subtypes and verify the authenticity of their differences and the association with tumor biological phenomena. Feature selection is the core of the S To M strategy. Over the years, there are 3 approaches to the selection of stratified features in the S To M strategy: data distribution-based, biological features, and multi-omics approaches. Unlike S To M strategy, M To M strategy is based on the concept of systems biology and presents a landscape of the differences and associations between different omics within different subtypes. The core step of the M To M strategy is data dimensionality reduction, which puts multi-omics data into a low-dimensional stacked matrix, providing input for the subsequent cluster analysis. Generally, M To M strategy stratification algorithms can be classified into three categories: similarity-based, integration-based, and deep learning. We believe that both S To M and M To M need to pay attention to the combination of data and the applicability and practicality of related software when using it for cancer subtype analysis. In the end, we summarized the current multi-omics characteristics of HCC subtypes. We found that HCC subtypes obtained by different methods may have a common feature, which suggests that more studies are needed in the future to summarize the more representative subclasses among them.

Key words hepatocellular carcinoma, multi-omics, subtype identification

DOI: 10.16476/j.pibb.2022.0397

* This study was supported by grants from The National Natural Science Foundation of China (61931013) and the Key Research and Development Program (2017YFC0111104).

** Corresponding author.

Tel: 86-15313254516, E-mail: lxq0811@bjut.edu.cn

Received: August 22, 2022 Accepted: October 7, 2022