

网络增强核融合方法的改进及其在乳头状肾细胞癌 多组学数据整合分子分型中的应用*

师国京¹ 李灵梅¹ 魏亿芳¹ 赵鑫¹ 房瑞玲^{1,2} 杨海涛³ 余红梅^{1,2} 张岩波^{1,2} 曹红艳^{1,3,Δ}

【摘要】目的 针对网络增强的相似网络融合(network enhancement fusion, ne-SNF)方法先融合不同组学网络,再对融合后的网络降噪,忽略了不同组学相似网络噪声对融合网络影响的问题,本文提出了改进的网络增强融合(improved network enhancement fusion, improved ne-SNF)方法,并探讨其在乳头状肾细胞癌(papillary renal cell carcinoma, PRCC)分子分型中的应用,识别 PRCC 高危患者,筛选重要通路及免疫浸润细胞。**方法** 通过模拟研究评估 improved ne-SNF 分型性能,并将其用于 PRCC 多组学数据的整合分型,利用 Cox 回归模型分析不同分型患者的预后风险;筛选不同分型的差异表达 mRNA (DEmRNAs)、miRNA (DEmiRNAs) 及差异甲基化基因(DMGs),并对其重合基因进行 KEGG 通路分析;最后对不同分型患者进行免疫细胞浸润分析。**结果** 模拟研究结果表明 improved ne-SNF 在不同信号比例和噪声强度下的分型准确性均优于 SNF 和 ne-SNF。improved ne-SNF 方法将 PRCC 患者分为高危组和低危组,高危组患者的死亡风险是低危组的 7.727 倍;筛选出 3511 个 DEmRNAs, 96 个 DEmiRNAs 及 3426 个 DMGs,其联合分析的 649 个重合基因得到 42 条有统计学差异的 KEGG 通路。此外,筛选出 3 种在不同分型中存在统计学差异的免疫浸润细胞。**结论** improved ne-SNF 分型性能优于 SNF 和 ne-SNF,且能够有效识别 PRCC 预后高风险患者,并筛选出 PRCC 重要通路及相关免疫浸润细胞,为 PRCC 的治疗及预后提供新的思路和参考依据。

【关键词】 改进的网络增强融合 乳头状肾细胞癌 多组学数据整合 分子亚型

【中图分类号】 R195.1 **【文献标识码】** A **DOI** 10.11783/j.issn.1002-3674.2024.03.010

An Improved Network – Enhanced Fusion Method and its Application in Papillary Renal Cell Carcinoma Subtyping using Multi-omics Data

Shi Guojing, Li Lingmei, Wei Yifang, et al (Department of Health Statistics, Public Health of School, Shanxi Medical University(030001), Taiyuan)

【Abstract】Objective The Network Enhancement Fusion(ne-SNF) method had no denoising strategy for the networks from each omics data type, which would weaken the power of network fusion. To address this problem, we proposed an improved Network Enhancement Fusion(improved ne-SNF) model, and further applied the method to subtype identification of papillary renal cell carcinoma. **Methods** We conducted simulation studies to compare the performances of the improved ne-SNF method with the ne-SNF and SNF method, and applied improved ne-SNF method to integrate multi-omics data of PRCC. Cox regression model was performed to evaluate the prognostic risk of different subtypes. Differentially expressed mRNAs (DEmRNAs), miRNAs (DEmiRNAs) and differentially methylated genes (DMGs) with different subtypes were screened. KEGG pathway analysis was performed for the overlapping genes of three gene sets. Finally, the immune cell infiltration analysis was performed for patients with different subtypes. **Results** The improved ne-SNF method outperformed both SNF and ne-SNF approach in various simulation scenarios. In subsequent subtyping application, PRCC patients were divided into high-risk and low-risk groups, and the risk of death was 7.727 times higher in the high-risk group than in the low-risk group. A total of 3511 DEmRNAs, 96 DEmiRNAs and 3426 DMGs were identified. Among them, 649 overlapping genes yielded 42 KEGG pathways with statistical differences. In addition, 3 immune filtering cells showed statistical significance. **Conclusion** The improved ne-SNF performed better than SNF and ne-SNF, and the identified subtypes of PRCC may provide important clues and basis for treatment of PRCC patient.

【Key words】 Improved ne-SNF; Papillary renal cell carcinoma; Multi-omics data integration; Molecular subtyping

乳头状肾细胞癌(papillary renal cell carcinoma, PRCC)占肾细胞癌的 10%~15%^[1],目前还没有针对 PRCC 的特异性治疗方法^[2],寻求其新型靶标极为重要。PRCC 是一种异质性疾病^[3],根据组织学特征分

为 Type I 和 Type II 两种亚型,但在临床上对同一组织亚型的 PRCC 患者采用相同的治疗方案,其治疗反应和预后结局却经常不同,究其原因,PRCC 的发生和发展涉及 mRNA、miRNA、DNA 甲基化等多个层面的改变和相互间的复杂调控机制,为实现 PRCC 患者的精准诊疗,需整合多个层面数据进行更精准的分型。随着测序技术的发展,积累了大量的多种组学数据,多组学数据整合分型的研究也随之兴起^[4]。整合多个组学的的数据可以同时捕捉到 PRCC 在不同组学上的关联和潜在的因果关系^[5],形成多种数据类型间的信息互补,从多层次揭示 PRCC 的复杂调控机制。因此利

* 基金项目:国家自然科学基金资助项目(71403156, 81872717, 82273742);山西省应用基础研究计划项目(202303021211130);山西医科大学博士启动基金项目(BS201722)

1.山西医科大学卫生统计教研室(030001)

2.重大疾病风险评估山西省重点实验室

3.河北医科大学流行病学与卫生统计学教研室

Δ通信作者:曹红艳,E-mail:caohy@sxmu.edu.cn

用 PRCC 多组学数据建立其分子分型,以此确定患者个体化治疗策略,是 PRCC 临床诊断和治疗的重要发展方向。

多组学数据整合存在不同组学间和组内基因间的交互作用及基因对肿瘤的复杂调控等非线性问题,基于核的方法,通过构建核函数将数据的原始特征映射到高维特征空间来解决非线性问题^[6]。相似网络融合(similarity network fusion, SNF)^[7],通过对每个组学应用核方法建立相似网络并采用信息传递^[8]的方法,将多个相似网络融合为一个相似网络,从而解决这一问题。然而,由于测量技术的限制和实验操作误差等,组学数据中存在大量的噪声,容易导致样本间存在虚假关联,弱化 SNF 分型性能^[9]。针对这一问题,Wu 等提出基于网络增强(network enhancement, NE)^[10]的 SNF,并将其命名为 ne-SNF^[11],通过网络增强策略^[10]消除 SNF 融合网络中由噪声引起的弱关联,增强样本之间的实际联接,提高了 SNF 的分型性能。但 ne-SNF 对融合后的网络降噪,忽略了不同组学相似网络噪声对融合网络的影响。

因此,本文提出改进的网络增强融合(improved network enhancement fusion, improved ne-SNF)方法,对每个组学相似网络分别采用网络增强^[10]策略,去除每个相似网络中由噪声引起的弱关联,然后利用信息传递^[8]方法,将多个降噪相似网络迭代融合为一个融合相似网络,最后采用谱聚类^[12]方法对融合网络进行样本分型。本文采用 improved ne-SNF 对 PRCC 患者 miRNA、mRNA 及 DNA 甲基化数据进行整合分型,针对不同分型,筛选差异基因并进行 KEGG 通路分析,研究不同分型免疫浸润细胞,为 PRCC 患者的预后判断和个体化治疗提供参考依据。

资料与方法

1. 数据来源与处理

利用 R 包 TCGAbiolinks^[13] 下载 PRCC 患者的 mRNA、miRNA、DNA 甲基化和临床数据,并进行数据预处理。首先,选择 Illumina 人源甲基化 450 芯片平台的注释文件,对启动子区域的 CpG 甲基化位点进行注释,启动子区域定义为转录起始位点 2kbp 内的区域^[14],并且剔除性染色体上的启动子区 CpG 位点;然后针对三种组学数据,删除缺失值的比例超过 30% 的样本并用 K 最邻近(K-nearest neighbor, KNN)^[15] 算法填补缺失值,同时对 mRNA 和 miRNA 进行 log₂ 转换;最后删除三种组学数据中小于剩余特征平均值第五百分位的特征。最终得到 207 例患者的 14921 个 mRNAs,394 个 miRNAs 及 44241 个启动子 CpG 甲基化位点。

2. improved ne-SNF 方法

针对 ne-SNF 方法^[11]先融合不同组学网络,再对融合后的网络降噪,忽略了不同组学相似网络噪声对融合网络的影响,我们提出了 improved ne-SNF 方法。Improved ne-SNF 方法在 SNF^[7]基础上,首先对每个组学数据构造相似网络,其次通过网络增强策略^[10]去除每个组学相似网络中由噪声引起的弱关联,然后利用信息传递^[8]的方法将多个组学的降噪相似网络迭代融合为一个融合相似网络,最后采用谱聚类^[12]方法对融合相似网络进行样本分型。技术路线如图 1。

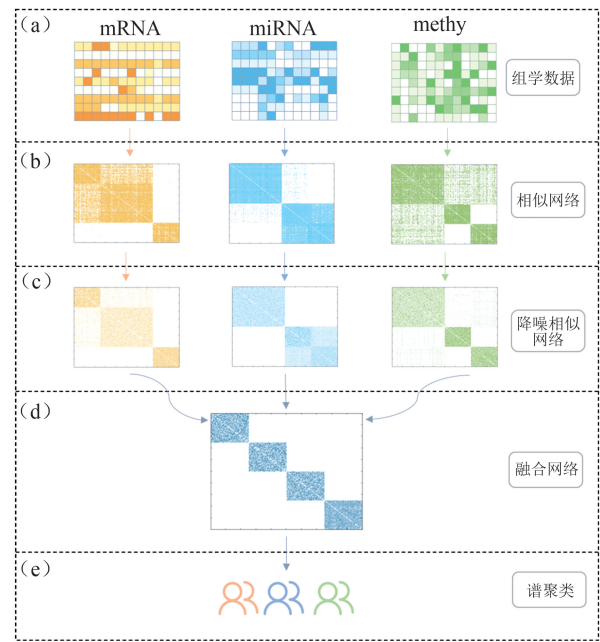


图 1 improved ne-SNF 技术路线

(1) 构建单个组学的相似网络

假定某组学数据有 n 个样本,其 $n \times n$ 的相似网络定义为:

$$A(i, j) = \exp\left(-\frac{d^2(x_i, x_j)}{\mu \varepsilon_{i,j}}\right)$$

其中 $\varepsilon_{i,j} = \frac{\text{mean}(d(x_i, N_i)) + \text{mean}(d(x_j, N_j)) + d(x_i, x_j)}{3}$,

$\text{mean}(d(x_i, N_i))$ 指样本 x_i 与其每个邻居间的距离的均值, N_i 与 N_j 分别表示样本 x_i 与样本 x_j 最近邻的集合。 $d^2(x_i, x_j)$ 指样本间的欧式距离, μ 一般设置为 0.5。

(2) 对单个组学的相似网络降噪

给定单个组学的相似网络矩阵 A , 采用网络增强^[10]的方法进行降噪, 构建对称双随机矩阵(doubly stochastic matrix, DSM) T , 步骤如下:

$$S_{ne}(i, j) = \frac{A(i, j)}{\sum_{k \in N_i} A(i, k)} \times I\{j \in N_i\}$$

$$T_{i,j} = \sum_{k=1}^n \frac{S_{ne}(i, k) S_{ne}(j, k)}{\sum_{v=1}^n S_{ne}(v, k)}$$

其中, $I_{i \in N_j}$ 为指示函数, N_i 为包括 x_i 在内的最近邻的集合, T 为编码原网络 A 局部结构的对称双随机矩阵。

进一步对 T 进行扩散, 定义为 $W_{t+1} = \alpha T \times W_t \times T + (1-\alpha)T$, t 表示迭代步骤, α 为正则化参数, 初始化 W_0 可以设置为 SNF 的相似网络 A , 在每次迭代中, W_t 能保证对称双随机性, 通过迭代收敛, 得到最终降噪相似网络 W_{ne} 。

(3) 多个降噪相似网络的融合

对单个组学的降噪相似网络 W_{ne} , 通过归一化得到全局相似网络 P , 同时采用 KNN 算法得到局部相似网络 Q 。对于第 m 个组学的局部相似网络 $Q^{(m)}$, 使用其余组学的全局相似网络 $P^{(k)}$ 迭代更新, 实现不同组学全局相似网络的融合, 得到第 m 个组学的相似网络 $P^{(m)}$, 最后对所有组学 (M) 的相似网络 $P^{(m)}$ 求均值得到最终的降噪融合网络 P_{final} 。

$$P(i, j) = \begin{cases} \frac{W_{ne}(i, j)}{2 \sum_{k \neq i} W_{ne}(i, k)}, & j \neq i \\ 1/2, & j = i \end{cases}$$

$$Q(i, j) = \begin{cases} \frac{W_{ne}(i, j)}{\sum_{k \in N_i} W_{ne}(i, k)}, & j \in N_i \\ 0, & otherwise \end{cases}$$

$$P^{(m)} = Q^{(m)} \times \frac{\sum_{k \neq m} P^{(k)}}{M-1} \times (Q^{(m)})^T, \quad m = 1, \dots, M$$

$$P_{final} = \frac{\sum_{m=1}^M P^{(m)}}{M}$$

(4) 聚类分型

对降噪融合网络 P_{final} 进行谱聚类^[12]分型, 最优分型数为使旋转代价^[7] (rotation cost) 最小的值。

3. 模拟研究

为了验证 improved ne-SNF 方法的分型优势, 进行了模拟研究。模拟产生三种组学数据, 每个组学数据包括预设的聚类结构和额外的噪声, 即 $X_{il} = mean^s + \varepsilon_{il}$, X_{il} 表示 200×1000 的矩阵, 其中行表示样本, 列表示特征; $\varepsilon_{il} \sim N(0, \sigma^2)$ 表示样本 i 中特征 l 的随机噪声, 其中 σ^2 为 1, 4, 9 分别表示低、中、高噪声; $mean^s$ 表示预设的聚类结构, 每个组学从 (0, 1, 2, 3) 中随机选取 3 个不重复的值作为这一组学每个亚型信号特征的基准值, 每个组学设置三个亚型, 而样本真实亚型设为 4 类, 即任一个组学均不能得到完整的分型, 只有整合三个组学数据才能确定最终的四个亚型。同时, 研究不同信号特征比例对分型的影响, 信号特征比例表示在所有特征中信号特征所占比例, 本研究中设置信号特征比例分别为 5%、10% 和 15%。

对不同的模拟数据集均进行 1000 次模拟, 并使用归一化互信息 (normalized mutual information,

NMI)^[16] 作为模拟研究分型结果的评价指标, NMI 定义为 $NMI(U, V) = \frac{I(U, V)}{\sqrt{H(U)H(V)}}$, 其中 $I(U, V)$ 是同

一数据分型结果 U 和 V 的互信息, $H(U)$ 表示分型 U 的熵。NMI 值范围在 0~1 之间, NMI 值越高, 表明分型结果与真实标签的一致性程度越高, 即分型准确性越好。

4. 基于 PRCC 分型结果的下游分析

(1) Cox 比例风险模型

针对 PRCC 的分型结果, 在控制病理分期、年龄和性别等协变量的情况下, 采用 Cox 比例风险模型对患者的分型结果进行预后评估。

(2) 筛选差异基因

通过 Kruskal-Wallis 秩和检验筛选差异表达的 mRNA (DEmRNAs)、miRNA (DEmiRNAs) 和差异甲基化基因 (DMGs), 阈值为 $P_{adj} < 0.01$; 并通过超几何分布检验筛选在各个亚型上富集的基因, 阈值为 $P_{adj} < 0.01$ 。

(3) KEGG 通路分析

通过 miRTarBase^[17] 数据库预测 DEmiRNAs 的靶基因, 然后采用 KOBAS^[18] 在线工具对 DEmRNAs 与 DEmiRNAs 靶基因和 DEmRNAs 与 DMGs 的重合基因进行 KEGG 通路分析。

(4) 免疫细胞浸润

使用 TIMER^[19] 中的 MCP-counter^[20] 算法估计肿瘤细胞成分, 并通过非参数检验选择在不同分型中存在差异的免疫浸润细胞, 阈值为 $P_{adj} < 0.05$ 。

结 果

1. 模拟研究结果

模拟研究得到的 NMI 均值和标准差结果如表 1, 随着信号比例的增加和噪声的降低, 不同方法的 NMI 均随之增加, 其中, improved ne-SNF 在不同信号比例和噪声强度下, NMI 值均高于 ne-SNF 和 SNF 方法。

表 1 模拟数据集上的 NMI 值 [mean(SD)]

信号比	方法	低噪声	中等噪声	高噪声
5%	improved ne-SNF	0.936(0.027)	0.619(0.039)	0.312(0.045)
	ne-SNF	0.841(0.052)	0.494(0.044)	0.282(0.038)
	SNF	0.799(0.052)	0.412(0.049)	0.272(0.039)
10%	improved ne-SNF	1.000(0.003)	0.853(0.039)	0.631(0.039)
	ne-SNF	0.988(0.014)	0.749(0.051)	0.510(0.045)
	SNF	0.954(0.062)	0.691(0.044)	0.429(0.050)
15%	improved ne-SNF	1.000(0.000)	0.959(0.023)	0.751(0.040)
	ne-SNF	0.996(0.008)	0.896(0.036)	0.662(0.042)
	SNF	0.890(0.092)	0.852(0.049)	0.592(0.046)

2. PRCC 分析结果

为验证 improved ne-SNF 方法的优越性, 将其与

SNF、ne-SNF 分别应用于 PRCC 多组学数据,并通过 log-rank 检验进行比较,结果如表 2。三种方法的最佳分型数均为 3,improved ne-SNF 的分型结果在生存时间上存在显著差异,且 P 值最小。生存曲线如图 2A,通过 log-rank 检验发现分组 1 与分组 2 生存率无统计学差异($\chi^2 = 0.358, P = 0.7$),因此,将分组 1 和分组 2 合并,命名为低危组,而分组 3 命名为高危组,生存曲线如图 2B。高危组与低危组的生存曲线分离程度高,分型的生存差异明显。两组基本资料如表 3。

表 2 PRCC 亚型分析

肿瘤类型 方法	PRCC		
	SNF	ne-SNF	improved ne-SNF
分型数目	3	3	3
log-rank p	7.03×10^{-6}	1.82×10^{-4}	2.87×10^{-6}

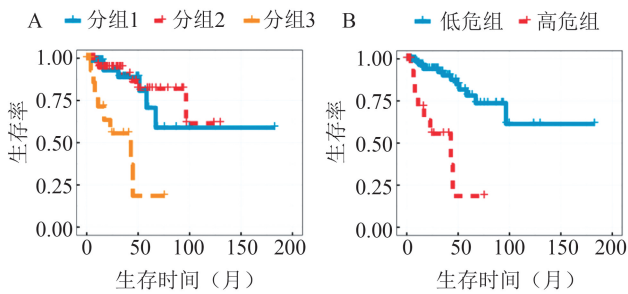


图 2 PRCC 分型生存曲线

表 3 PRCC 患者不同分型的基本资料 [n(%)]

项目	低危组	高危组
例数	186(89.86)	21(10.14)
年龄($\bar{x} \pm s$, 岁)	61.38 \pm 10.41	55.67 \pm 15.59
性别		
女性	45(24.19)	15(71.43)
男性	141(75.81)	6(28.57)
种族		
白人	141(75.81)	14(66.67)
黑人或非裔	42(22.58)	6(28.57)
亚洲人	3(1.61)	1(4.76)
病理分期		
I 期	130(69.89)	8(38.10)
II 期	13(6.99)	1(4.76)
III 期	35(18.82)	8(38.10)
IV 期	8(4.30)	4(19.05)
生存状态		
存活	172(92.47)	13(61.90)
死亡	14(7.53)	8(38.10)

在校正年龄、性别及病理分期协变量的情况下,拟合 Cox 回归模型研究不同亚型与预后的关系,结果如表 4,高危组患者的死亡风险是低危组的 7.727 倍。与此同时,病理分期是 PRCC 患者的死亡危险因素,IV 期患者的死亡风险是 I 期患者的 15.247 倍。

表 4 PRCC 患者 Cox 回归分析结果

变量	b(S.E)	Z	P	HR(95% CI)
分型				
高危组	2.045(0.496)	4.120	3.79×10^{-5}	7.727(2.921~20.438)
年龄	0.011(0.021)	0.501	0.616	1.011(0.969~1.054)
性别	0.034(0.491)	0.070	0.945	1.035(0.395~2.710)
病理分期				
II 期	0.460(1.086)	0.424	0.672	1.585(0.189~13.322)
III 期	0.824(0.546)	1.509	0.131	2.280(0.782~6.649)
IV 期	2.724(0.571)	4.772	1.82×10^{-6}	15.247(4.980~46.679)

*: 低危组为分型的参考组, I 期为病理分期的参考组。

3. 筛选差异基因

差异表达分析共筛选出 96 个 DEmiRNAs, 其中 6 个上调, 90 个下调; 3511 个 DEmRNAs 中上调 1924 个, 下调 1587 个; 5767 个差异启动子 CpG 甲基化位点映射到 3426 个 DMGs, 其中 1307 个高甲基化基因, 2119 个低甲基化基因。DEmiRNAs, DEmRNAs 和 DMGs 的热图如图 3, 差异表达基因在不同分组的分布明显不同。

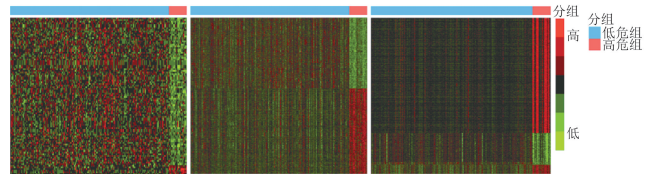


图 3 差异表达基因热图

4. KEGG 通路分析

对 96 个 DEmiRNAs 进行 MiRTarBase^[17] 预测得到 488 个靶基因。通过对 488 个 DEmiRNAs 靶基因, 3511 个 DEmRNAs 及 3426 个 DMGs 联合分析得到三个组学的 14 个重合基因, 其中 CACNA1C、PACS1、PYCR1、CENPN 和 HLX 高表达, 且低甲基化, 对应的 miRNAs 低表达。与此同时, 84 个基因既 mRNA 差异表达又 miRNAs 差异表达, 551 个基因出现差异表达同时发生了异常甲基化, 如图 4。

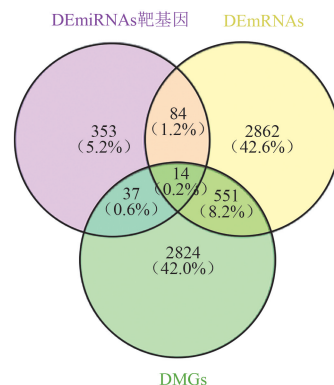


图 4 差异基因韦恩图

对 649 个重合基因进行 KEGG 通路分析, 共筛选出 42 条 KEGG 通路, 基因显著富集的前 10 条通路如

图 5,表明不同分型的差异基因与 PRCC 存在密切相关的生物作用。

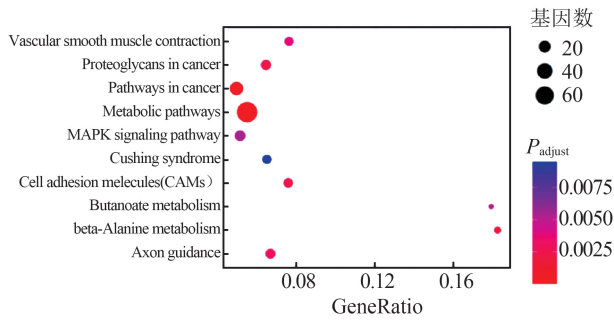


图 5 KEGG 通路分析

5. 免疫细胞浸润

单核细胞系细胞、癌相关成纤维细胞与内皮细胞

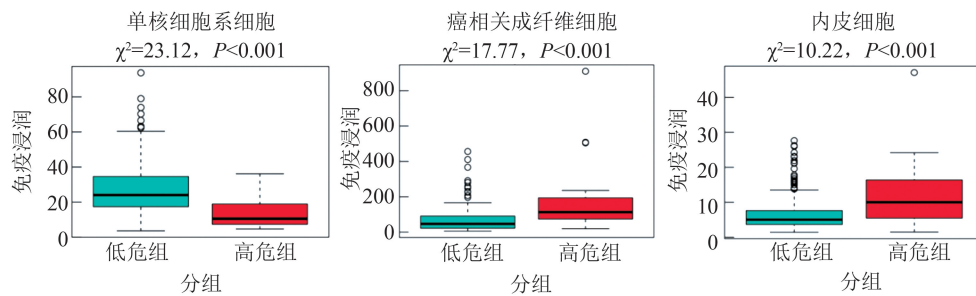


图 6 免疫细胞浸润在不同分型的差异

高危组和低危组的差异分子靶标,同时受到 miRNA、mRNA 和 DNA 甲基化调控的 14 个基因中, *CACNA1C*、*PACS1*、*PYCR1*、*CENPN* 和 *HLX* 高表达,且低甲基化,对应的 miRNAs 低表达,值得进一步研究。已有研究证实高表达的 *PYCR1* 通过 Akt/mTOR 途径调控 PRCC 的进展,*PYCR1* 可作为未来临床 PRCC 治疗的潜在靶点^[21]。*CACNA1C* 与子宫内膜癌患者的总生存时间相关^[22]。*CENPN* 高表达是肝细胞癌患者不良预后的独立危险因素,*CENPN* 作为肝癌的致癌基因,可能是肝癌的治疗靶点和预后指标^[23]。*CACNA1C* 和 *CENPN* 与 PRCC 的关系还需进一步验证。

高危组和低危组的差异表达基因分布明显不同,对其重合基因通过 KEGG 通路分析进行生物功能注释。基因显著富集的前 10 条通路中与癌症直接相关的通路有 2 条,为癌症通路和癌症中的蛋白多糖。其余癌症通路中, β -丙氨酸信号通路的激活与肾细胞癌低生存率存在关联^[24];细胞粘附分子与多种恶性肿瘤的进展相关^[25],如 L1 细胞粘附分子是透明细胞肾细胞癌的重要预后因素^[26];MAPK 信号通路在肾细胞癌的增殖、迁移和侵袭中发挥关键作用^[27];轴突导向通路中 *Slit2* 可能是肾细胞癌中潜在的肿瘤抑制因子^[28]。现有研究已表明这四条通路与肾细胞癌的进展有关,其与 PRCC 的因果关系还需进一步验证。

本研究筛选出 3 种在高危组和低危组中存在显著

在 高危组与低危组中存在显著差异(图 6),其中,单核细胞系细胞在 高危组中的丰度显著低于低危组,而癌相关成纤维细胞与内皮细胞在 高危组中的丰度显著高于低危组。

讨论

本文针对 ne-SNF 方法忽略了不同组学相似网络噪声对融合网络的影响,提出了 improved ne-SNF 方法,并将其应用于 PRCC 患者多组学数据的整合分型。improved ne-SNF 在模拟研究和 PRCC 分型应用上均优于 SNF 和 ne-SNF。improved ne-SNF 方法将 PRCC 患者分为高危组和低危组,高危组患者的死亡风险是低危组的 7.727 倍,为 PRCC 患者开展针对性治疗提供了重要理论依据。

差异的免疫细胞,其中单核细胞系细胞在 高危组中的丰度显著低于低危组,而癌相关成纤维细胞与内皮细胞在 高危组中的丰度显著高于低危组。现有研究中晚期肾癌的一线治疗药物舒尼替尼通过拮抗血管内皮生长因子等受体抑制肿瘤生长^[29],表明高浸润的内皮细胞与促进 PRCC 的进展可能存在相关关系;已有研究表明癌相关成纤维细胞通过促进透明细胞肾细胞癌肿瘤细胞的增殖,降低患者的生存时间,并使靶向治疗产生抗性^[30],表明高浸润的癌相关成纤维细胞与 PRCC 较差的预后可能存在潜在关系。新发现的免疫细胞如内皮细胞、癌相关成纤维细胞等有可能为改善 PRCC 患者的生存提供新的临床药物靶点。

总之,本文基于 ne-SNF 的框架提出了 improved ne-SNF 方法,有效地提高了分型准确性,为多组学数据整合分析提供了新的方法。在对 PRCC 分子分型研究中,识别出了 PRCC 预后高危患者,筛选出在不同分型中存在显著差异的免疫细胞及与 PRCC 进展相关的通路,可作为免疫治疗及干扰或促进相关通路的参考,为 PRCC 患者的靶向治疗提供新的思路和重要的参考依据。

参考文献

[1] Shuch B, Amin A, Armstrong AJ, et al. Understanding pathologic variants of renal cell carcinoma: distilling therapeutic opportunities

- from biologic complexity. *European Urology*, 2015, 67(1):85-97.
- [2] Singer EA, Bratslavsky G, Linehan WM, et al. Targeted therapies for non-clear renal cell carcinoma. *Targeted Oncology*, 2010, 5(2):119-129.
- [3] Cancer Genome Atlas Research Network. Comprehensive molecular characterization of papillary renal-cell carcinoma. *New England Journal of Medicine*, 2016, 374(2):135-145.
- [4] 徐臻旖, 王策, 侯艳, 等. 多组学数据分析中关联网络融合 ANF 算法及应用研究. *中国卫生统计*, 2020, 37(6):6.
- [5] Lock EF, Hoadley KA, Marron JS, et al. Joint and individual variation explained (JIVE) for integrated analysis of multiple data types. *The Annals of Applied Statistics*, 2013, 7(1):523.
- [6] Cristianini N, Shawe-Taylor J. *Kernel methods for pattern analysis*. Cambridge: Cambridge University Press, 2004.
- [7] Wang B, Mezlini AM, Demir F, et al. Similarity network fusion for aggregating data types on a genomic scale. *Nat Methods*, 2014, 11(3):333-337.
- [8] Pearl J. *Probabilistic reasoning in intelligent systems; networks of plausible inference*. Morgan kaufmann, 1988.
- [9] Ji J, Zhang A, Liu C, et al. Survey: Functional module detection from protein-protein interaction networks. *IEEE Transactions on Knowledge and Data Engineering*, 2012, 26(2):261-277.
- [10] Wang B, Pourshafeie A, Zitnik M, et al. Network enhancement as a general method to denoise weighted biological networks. *Nature Communications*, 2018, 9(1):1-8.
- [11] Wu Y, Wang H, Li Z, et al. Subtypes identification on heart failure with preserved ejection fraction via network enhancement fusion using multi-omics data. *Computational and Structural Biotechnology Journal*, 2021, 19:1567-1578.
- [12] Ng AY, Jordan MI, Weiss Y. On spectral clustering: Analysis and an algorithm. *NIPS*, 2002:849-856.
- [13] Colaprico A, Silva TC, Olsen C, et al. TCGAAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res*, 2016, 44(8):e71.
- [14] Gusev A, Lee SH, Trynka G, et al. Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. *The American Journal of Human Genetics*, 2014, 95(5):535-552.
- [15] Hastie T, Tibshirani R. Discriminant adaptive nearest neighbor classification. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 1996, 18(6):607-616.
- [16] Kato H, Ichinose Y, Ohta M, et al. A randomized trial of adjuvant chemotherapy with uracil-tegafur for adenocarcinoma of the lung. *New England Journal of Medicine*, 2004, 350(17):1713-1721.
- [17] Chou CH, Shrestha S, Yang CD, et al. miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic Acids Research*, 2018, 46(D1):D296-D302.
- [18] Xie C, Mao X, Huang J, et al. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic acids Research*, 2011, 39(suppl 1_2):W316-W322.
- [19] Li B, Severson E, Pignon JC, et al. Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. *Genome Biology*, 2016, 17(1):1-16.
- [20] Becht E, Giraldo NA, Lacroix L, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biology*, 2016, 17(1):1-20.
- [21] Wang QL, Liu L. PYCR1 is associated with papillary renal cell carcinoma progression. *Open Medicine*, 2019, 14(1):586-592.
- [22] Qiao Z, Jiang Y, Wang L, et al. Mutations in KIAA1109, CACNA1C, BSN, AKAP13, CELSR2, and HELZ2 are associated with the prognosis in endometrial cancer. *Frontiers in Genetics*, 2019, 10:909.
- [23] Wang Q, Yu X, Zheng Z, et al. Centromere protein N may be a novel malignant prognostic biomarker for hepatocellular carcinoma. *PeerJ*, 2021, 9:e11342.
- [24] Popławski P, Tohge T, Bogusawska J, et al. Integrated transcriptomic and metabolomic analysis shows that disturbances in metabolism of tumor cells contribute to poor survival of RCC patients. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, 2017, 1863(3):744-752.
- [25] Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell*, 2000, 100(1):57-70.
- [26] Allory Y, Matsuoka Y, Bazille C, et al. The L1 cell adhesion molecule is induced in renal cancer cells and correlates with metastasis in clear cell carcinomas. *Clinical Cancer Research*, 2005, 11(3):1190-1197.
- [27] Hong B, Zhou J, Ma K, et al. TRIB3 promotes the proliferation and invasion of renal cell carcinoma cells via activating MAPK signaling pathway. *International Journal of Biological Sciences*, 2019, 15(3):587.
- [28] Ma WJ, Zhou Y, Lu D, et al. Reduced expression of Slit2 in renal cell carcinoma. *Medical Oncology*, 2014, 31(1):1-7.
- [29] Motzer RJ, Hutson TE, Tomczak P, et al. Overall survival and updated results for sunitinib compared with interferon alfa in patients with metastatic renal cell carcinoma. *Journal of Clinical Oncology*, 2009, 27(22):3584.
- [30] Chakiryan NH, Kimmel GJ, Kim Y, et al. Geospatial cellular distribution of cancer-associated fibroblasts significantly impacts clinical outcomes in metastatic clear cell renal cell carcinoma. *Cancers*, 2021, 13(15):3743.

(责任编辑:郭海强)