

· 论著 ·

## 基于生存结局加权多组学数据整合的胶质瘤分子分型\*

贾聪聪<sup>1</sup> 杜港<sup>1</sup> 赵鑫<sup>1</sup> 师国京<sup>1</sup> 房瑞玲<sup>1,2,3</sup> 李治<sup>4</sup> 张岩波<sup>1,2,3</sup> 曹红艳<sup>1,2,3,Δ</sup>

**【摘要】**目的 探讨将生存结局加权的组学数据整合方法 survClust 应用于胶质瘤 (glioma) 数据,以识别具有显著分子异质性和预后差异的胶质瘤分子分型。方法 采用 survClust 方法对中国胶质瘤基因数据库 (Chinese glioma genome atlas, CGGA) 的胶质瘤多组学数据进行结局加权整合分型,并拟合 Cox 比例风险模型评估不同分型患者预后。对不同分型间的差异表达基因 (DEmiRNAs, DEmRNAs, DMGs) 进行筛选,对 DEmiRNAs 靶基因、DEmRNAs、DMGs 的重合基因进行 GO 功能注释;最后对不同分型患者进行免疫浸润分析。结果 survClust 将胶质瘤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 2.931 倍。不同分型差异基因的分布存在差异,共筛选出 194 个 DEmiRNAs, 3396 个 DEmRNAs, 1230 个 DMGs。将 189 个重合基因进行 GO 功能注释,得到 52 条差异具有统计学意义的 GO 生物项。此外,不同分型在 B 淋巴细胞、CD4+T 细胞、CD8+T 细胞、中性粒细胞、巨噬细胞和髓样树突状细胞的免疫浸润水平存在统计学差异。结论 结局加权整合算法 survClust 能够有效识别兼具分子异质性和显著预后差异的胶质瘤亚型,同时基于分型结果筛选出的潜在生物标志物将为胶质瘤的个性化治疗提供科学理论依据。

**【关键词】** 生存结局加权聚类 分子亚型 多组学数据整合 胶质瘤

**【中图分类号】** R195.1 **【文献标识码】** A **DOI** 10.11783/j.issn.1002-3674.2024.05.002

## Molecular Subtyping of Gliomas based on the Integration of Survival Outcome-weighted Multi-omics Data

Jia Congcong, Du Gang, Zhao Xin, et al (Division of Health Statistics, Shanxi Medical University, Taiyuan 030001)

**【Abstract】 Objective** To investigate the survival outcome-weighted method, survClust, and apply it to the multi-omics data of gliomas, to identify molecular subtypes of gliomas with significant molecular heterogeneity and prognostic differences. **Methods** The multi-omics data from the Chinese Glioma Genome Atlas (CGGA) was used for outcome weighted integrated subtyping using the survClust method, and a Cox proportional risk model was fitted to assess the prognosis of patients with different subtypes. Differentially expressed genes (DEmiRNAs, DEmRNAs, DMGs) between different subtypes were screened, and Gene Ontology (GO) analysis was performed for overlapping genes among DEmiRNAs target genes, DEmRNAs, and DMGs. Finally, we performed immune infiltration analysis between different subtypes. **Results** The patients with gliomas were divided into high-risk and low-risk groups by using survClust, and the risk of death for patients in the high-risk group was 2.931 times higher than in the low-risk group. The distribution of differential genes was significantly different among different subtypes, and 194 DEmiRNAs, 3396 DEmRNAs, and 1230 DMGs were screened. 189 overlapping genes were used for GO analysis, and 52 GO terms with statistically significant differences were obtained. In addition, the level of immune infiltration differed statistically between the different subtypes in terms of B cells, CD4+ T cells, CD8+ T cells, neutrophils, macrophages and myeloid dendritic cells. **Conclusion** The outcome-weighted integration algorithm survClust can effectively identify subtypes of gliomas with both molecular heterogeneity and significant prognostic differences. At the same time, the potential biomarkers screened based on subtypes will provide a scientific and theoretical basis for individualized treatment of gliomas.

**【Key words】** Outcome-weighted clustering; Molecular subtype; Multi-omics integration; Gliomas

胶质瘤作为最常见的中枢神经系统肿瘤,年发病率达到了 3/10 万~6.4/10 万,在所有中枢神经系统肿瘤中的占比约为 23.3%<sup>[1]</sup>。根据世界卫生组织《WHO 中枢神经系统肿瘤分类指南》<sup>[2]</sup>,胶质瘤分为 WHO I 级至 IV 级,其中胶质母细胞瘤(胶质瘤 WHO IV 级)的治疗难度已经超越胰腺癌和肝癌,成为第一难治肿

瘤<sup>[3]</sup>。当前临床治疗中,主要依据组织病理学检验结果,对处于同一病理等级或症状相似的胶质瘤患者采用相同的治疗方式,治疗效果往往大相径庭,患者个体间具有高度的异质性,其多变的临床行为以及预后差异凸显出当前治疗体系的局限性<sup>[4-5]</sup>。因此,挖掘胶质瘤的潜在亚型,准确识别胶质瘤预后高风险患者,为胶质瘤患者的治疗策略提供新的靶点和个性化治疗方案,对于推动胶质瘤诊断治疗具有重要理论与现实意义。

肿瘤发展机制复杂,涉及基因变异、遗传、基因表达水平异常等多个层面。随着高通量测序技术的发展,获取了大量的高维基因组学数据<sup>[6]</sup>,如 CGGA 数据库<sup>[7]</sup>,收录了近 2000 例中国胶质瘤患者的基因组学数据与临床数据,为胶质瘤生物学研究提供了现实依据。针对胶质瘤开展多组学数据分型研究,能够挖掘

\* 基金项目:国家自然科学基金资助项目(82473739);山西省基础研究计划(202303021211130);山西省回国留学人员科研资助项目(2024-081);山西省高等教育“百亿工程”科技引导专项;山西省基础研究计划(20210302124409);山西省研究生科研创新项目(2023KY403)

1. 山西医科大学卫生统计教研室(030001)

2. 山西医科大学公共卫生学院重大疾病风险评估山西省重点实验室

3. 煤炭环境致病与防治教育部重点实验室

4. 中北大学运动人体科学教研室

Δ通信作者:曹红艳,E-mail: caohy@sxmu.edu.cn

不同数据类型间的潜在信息和因果关系<sup>[8]</sup>,从多层次揭示胶质瘤的调控机制<sup>[9]</sup>,提高分型的准确性。

患者预后结局是评价治疗效果的重要指标,然而现有的多组学整合分型方法大多为无监督方法,分型后再评价分子分型和生存结局之间的关系<sup>[10]</sup>,将分型和预后生存分析作为两个独立的过程,相互割裂<sup>[11]</sup>。针对这一局限,Arora<sup>[12]</sup>等提出了生存结局加权的整合分型方法 survClust,通过构建与生存结局相关的距离矩阵,同时采用  $K$  折交叉验证,避免在模型构建和预后评价时均使用生存数据而引起的“过拟合”,在识别显著预后差异的分子分型上独具优势。因此,本文采用 survClust 算法,整合胶质瘤患者的 mRNA、DNA 甲基化及 miRNA 数据进行生存结局加权分型,并研究不同分型的差异基因以及免疫浸润水平,为实现胶质瘤患者的精准诊疗提供参考依据。

## 数据与方法

### 1. 数据来源与处理

本研究从 CGGA 数据库收录的胶质瘤患者数据中筛选出同时具有三个组学 (mRNA 表达数据、miRNA 表达数据、DNA 甲基化数据) 的样本,同时删除在总生存期、生存状况、年龄、性别以及 WHO 分级等特征上存在缺失的样本,最终得到 110 例样本。然后分别按照均值、变异系数、方差等指标对三个组学数据矩阵的特征进行“过滤式”筛选:①计算不同组学特征均值,按百分位数保留均值较大的特征 (miRNA  $\geq$  特征均值 10 百分位数, mRNA 和甲基化  $\geq$  特征均值 15 百分位数);②进一步计算不同组学的特征变异系数,按百分位数保留变异系数较大的特征 (miRNA  $\geq$  特征变异系数 10 百分位数, mRNA 和甲基化  $\geq$  特征变异系数 15 百分位数);③最后计算不同组学特征方差,按百分位数保留方差较大的特征 (mRNA  $\geq$  特征方差 30 百分位数,甲基化  $\geq$  特征方差 15 百分位数)。最终得到 671 个 miRNAs, 9819 个 mRNAs 以及 8889 个以基因为单位的 DNA 甲基化。

### 2. 分析方法

#### (1) survClust 方法

生存结局加权的多组学数据整合分型方法 survClust,具体实现步骤如下:

#### ①计算加权距离矩阵

设数据矩阵  $X$  为  $n * p$  矩阵,其中  $n$  为样本数, $p$  为特征数,对每个特征拟合单因素 cox 比例风险模型以获取特征权重, $h(t) = h_0 \times \exp(x_j^T * \beta)$ ,  $t$  为生存时间, $x_j$  为矩阵  $X$  的第  $j$  列, $h_0$  是基准危险函数, $\beta$  为回归系数, $\exp(\beta)$  为风险比 (hazard ratio, HR)。取 HR 对数的绝对值即  $|\log HR|$  作为特征权重  $w$ ,当  $HR = 1$  时,对应  $|\log HR|$  为 0,则认为该特征与生存无关。 $W =$

$\text{diag}\{w_1, w_2, \dots, w_p\}$ , 矩阵  $X$  中的任意两个样本  $a$  和  $b$  间的加权距离  $d_w$  计算如下:

$$d_w(a, b) = \sqrt{(a-b)^T W (a-b)}$$

其中, $a$  和  $b$  分别是样本  $a$  和  $b$  的特征向量, $d_w$  值越小,两样本相似度越高。

进一步对数据进行归一化和标准化处理: $X' = X * W^{\frac{1}{2}}$  (归一化),  $\frac{X'}{\sum_i \sum_j x'_{ij}}$  (标准化),则样本  $a$  和  $b$  加权距离的计算转换为:

$$d_w(a', b') = d_w(b', a') = \sqrt{\sum_{j=1}^p (a'_j - b'_j)^2}$$

得到样本两两之间的加权距离矩阵  $D$ 。

#### ②整合加权距离矩阵

设有  $M$  个数据矩阵, $D_m$  为第  $m$  个 ( $m = 1, \dots, M$ ) 加权距离矩阵,对  $M$  个加权距离矩阵进行整合, $\gamma_m = \frac{1}{M} \forall m$ ,得到融合加权距离矩阵  $I_w$ :

$$I_w = \sum_{m=1}^M \gamma_m D_m$$

#### ③基于融合加权距离矩阵的聚类分型

在融合距离矩阵  $I_w$  的基础上,采用多维尺度变换 (multidimensional scaling, MDS)<sup>[13]</sup> 进行维度转换,进而使用  $K$ -means<sup>[14]</sup> 方法对样本进行聚类分型,依据 log-rank 检验统计量  $\chi^2$ 、标准化类内距离平方和 (standardized within-cluster sum of squares)<sup>[15]</sup> 来确定聚类分型数  $k$  的选择,选取  $\chi^2$  值最大以及类内差异尽可能小对应的分型结果为最佳分型。

#### ④ $K$ 折交叉验证

survClust 整合了患者生存结局信息,为避免预后评价时产生“过拟合”,并提高模型的泛化性能,在模型构建时引入了交叉验证。设置交叉验证的折数“ $K$ ”应结合实际分析样本量<sup>[16]</sup>,由于样本间的生存数据存在一定差异, $K$  值的选取将直接影响不同分区内生存数据的分布,从而进一步影响样本标签计算的准确性和稳定性,应根据实际情况避免分区内的样本量过多或过少。本研究中包含 110 例样本,规模较小,因此将“ $K$ ”设置为 3 开展交叉验证:a.将样本随机分成三个样本量均衡的分区,抽取其中一个分区作为测试数据集,另外两个为训练数据集;b.采用 survClust 方法获取训练数据集类别标签,测试数据集通过自身的分子特征向量以及由训练集标签结果提供的特征权重获取标签,未使用生存数据;c.进行 3 次不重复抽取,并计算 3 折交叉验证中所有测试集样本的类别标签;d.对所有的测试集标签采用质心重设标签方法,计算每个测试集的质心,在质心向量上的基础上,运行  $K$ -means,得到所有测试集样本的最新类别标签,并以此作为数据集的全部标签。对交叉验证过程开展多次 (如 50 次),对所有轮次的标签结果进行一致性投票,

选取次数最多的标签作为最终标签。

基于生存结局加权多组学数据整合的胶质瘤分子分型通过 R 软件包 survClust 实现。

### (2) 生存分析

基于 survClust 的分型结果,对不同分型患者的相关临床资料进行统计描述,定量资料采用均数±标准差( $\bar{x}\pm s$ )表示,分类资料采用构成比( $n, \%$ )表示。在校正病理分期、年龄、性别等协变量的基础上,采用 Schoenfeld 残差法检验协变量是否满足风险比例(proportional hazards, PH)假定,若满足直接拟合 Cox 比例风险模型进一步研究胶质瘤的不同分型和预后的关系,反之采用 Time-Varying Cox (TVC)模型分析。

### (3) 差异基因分析

采用 Kruskal-Wallis 秩和检验筛选出不同分型间的差异特征( $P_{adj} < 0.05$ ),进一步采用超几何分布检验筛选出在每个分型上富集的差异特征( $P_{adj} < 0.05$ ),得到最终的 DE miRNAs、DE mRNAs 和 DMGs。

### (4) GO 分析

使用 miRWalk 在线工具<sup>[17]</sup>(<http://mirwalk.umm.uni-heidelberg.de/>)预测 DE miRNAs 的靶基因,对 DE miRNAs 靶基因、DE mRNAs 以及 DMGs 进行联合分析,得到重合基因,并采用 clusterProfile<sup>[18]</sup>R 包对重合基因进行 GO<sup>[19]</sup>分析,以 $P_{adj} < 0.05$ 为阈值。

### (5) 免疫细胞浸润分析

使用 TIMER2.0 数据库<sup>[20]</sup>(<http://timer.cistrome.org/>)中的 TIMER 算法估计胶质瘤的细胞成分,通过 Kruskal-Wallis 秩和检验方法确定在不同分型中存在差异的免疫浸润细胞,阈值设定为 $P_{adj} < 0.05$ 。

## 结果

### 1. 胶质瘤患者分型结果

采用 survClust 方法对 110 名胶质瘤患者进行多组学数据整合分型,最佳分型数为 2 型, Kaplan-Meier 曲线如图 1 所示, log-rank 检验显示不同分型患者的生存率有显著统计学差异( $\chi^2 = 36.3, P < 0.001$ )。将生存预后较好的一组命名为“低危组”,生存预后较差的组别命名为“高危组”。不同分型胶质瘤患者的基本资料如表 1 所示。

表 1 不同分型胶质瘤患者的基本资料 [ $n(\%)$ ]

项目	类别	低危组	高危组
例数( $n, \%$ )		59(53.64)	51(46.36)
年龄( $\bar{x}\pm s$ )		37.66±9.3	41.02±14.1
性别( $n, \%$ )	男性	32(54.24)	30(58.82)
	女性	27(45.76)	21(41.18)
病理分期( $n, \%$ )	WHO II	50(84.75)	2(3.92)
	WHO III	7(11.86)	27(52.94)
	WHO IV	2(3.39)	22(43.14)
生存状态( $n, \%$ )	存活	36(61.02)	10(19.61)
	死亡	23(38.98)	41(80.39)

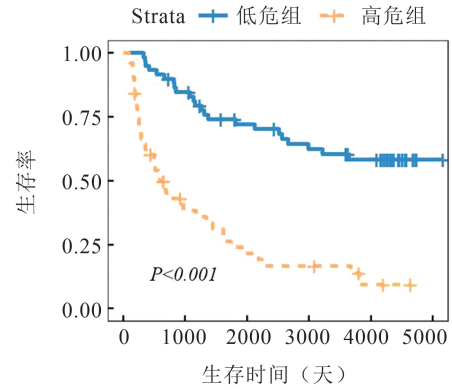


图 1 不同分型的 Kaplan-Meier 曲线

在校正年龄、性别、病理分期等协变量的情况下,采用 Schoenfeld 残差法对其进行检验,结果显示 Global Schoenfeld Test 对应  $P = 0.574$ ,即满足风险比例模型假定,进一步拟合 Cox 风险比例模型,研究不同分型间的预后,结果如表 2 所示,预后较差的高危组患者的死亡风险是低危组的 2.931 倍,同时“年龄”、“病理分期”对预后的影响具有统计学意义( $P < 0.001$ ),其中 WHO IV 级患者的死亡风险是 WHO II 级和 WHO III 级患者的 2.494 倍。

表 2 110 例胶质瘤患者多变量 Cox 回归分析结果

变量	$b(S.E)$	Z	P	HR(95% CI)
分型				
高危组	1.075(0.322)	3.342	<0.001	2.931(1.560~5.507)
性别	-0.163(0.256)	-0.637	0.524	0.849(0.514~1.403)
年龄	0.036(0.010)	3.461	<0.001	1.037(1.016~1.058)
病理分期				
WHO IV	0.914(0.318)	2.873	<0.001	2.494(1.337~4.652)

注:分型以低危组为对照;性别以男性为对照;病理分期以 WHO II 级和 WHO III 级为对照。

### 2. 差异基因分析结果

差异表达分析分别筛选出 194 个 DE miRNAs(其中上调 174 个,下调 20 个)、3396 个 DE mRNAs(其中上调 2040 个,下调 1356 个)以及 1230 个 DMGs(其中 800 个高甲基化基因,430 个低甲基化基因)。差异基因热图(图 2)显示高危组和低危组的差异特征分布明显不同。

使用“miRWalk”在线工具预测出 194 个 DE miRNA 的 61 个靶基因,对 DE mRNAs、DMGs 以及 DE miRNAs 靶基因进行联合分析,结果如图 3 所示。其中 7 个基因的差异表达可能受 DE miRNAs 调控,6 个基因的异常甲基化可能受到 DE miRNAs 调控,178 个基因既存在差异表达又发生了异常甲基化。

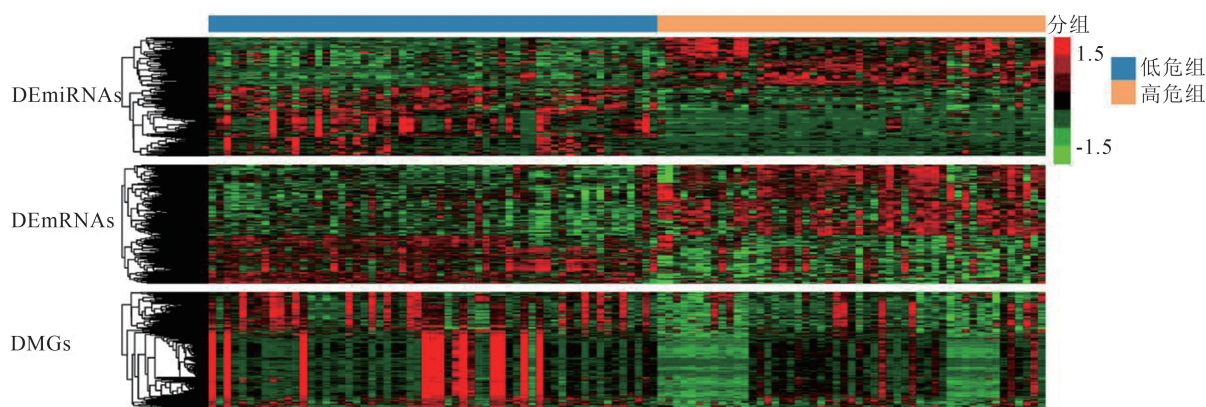


图2 不同分型差异表达基因热图

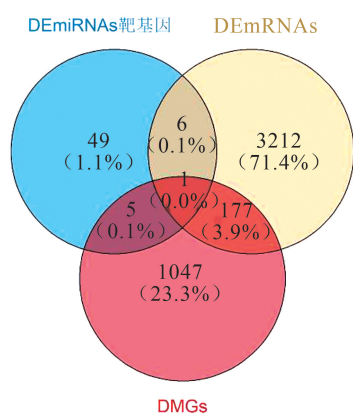


图3 差异基因韦恩图

### 3. GO 分析结果

对重合分析得到的 189 个基因进行 GO 分析,从分子功能、生物过程和细胞组成三个方面对基因和基因产物进行功能注释<sup>[19]</sup>,筛选得到 52 条差异具有统计学意义的 GO 生物功能项,其中差异最显著的前 10 项见图 4。

### 4. 免疫细胞浸润分析结果

免疫浸润分析结果表明,不同分型在 B 淋巴细胞、CD4+T 细胞、CD8+T 细胞、中性粒细胞、巨噬细胞和髓样树突状细胞的免疫浸润水平存在统计学差异,其中,髓样树突状细胞在高危组的丰度水平显著高于低危组(图 5)。

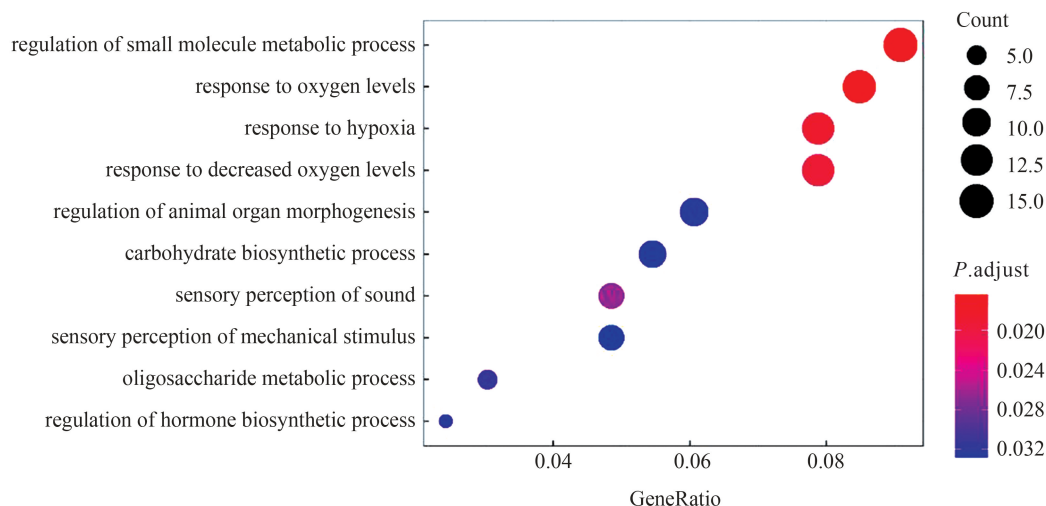


图4 GO 分析气泡图

## 讨论

本文采用 survClust 方法对胶质瘤多组学数据进行整合分型,将胶质瘤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 2.931 倍,不同分型在基因表达以及免疫浸润水平等方面均有显著差异,识别出了差异基因以及重要的免疫浸润细胞,实现了中国胶质瘤患者的分子分型。

重合的差异基因中, *KIT* 基因可能同时受到了

miRNA、mRNA 以及 DNA 甲基化的调控,研究表明 *KIT* 基因可通过抑制 STAT3 活性来诱导抗肿瘤巨噬细胞<sup>[21-22]</sup>,其表达水平降低会减少 F4/80+巨噬细胞的募集及其分泌因子的活性<sup>[21,23]</sup>,因此,高危组的不良预后可能与该基因的表达水平下调有关。GO 分析中, *PPARGC1A* 基因富集于多个具有统计学差异的生物项, *PPARGC1A* 基因转录率提高会促进 MTORC1 通路的激活,导致细胞的耗氧量增加,推动恶性胶质瘤细胞易受缺氧和营养流失,这有利于胶质母细胞瘤患

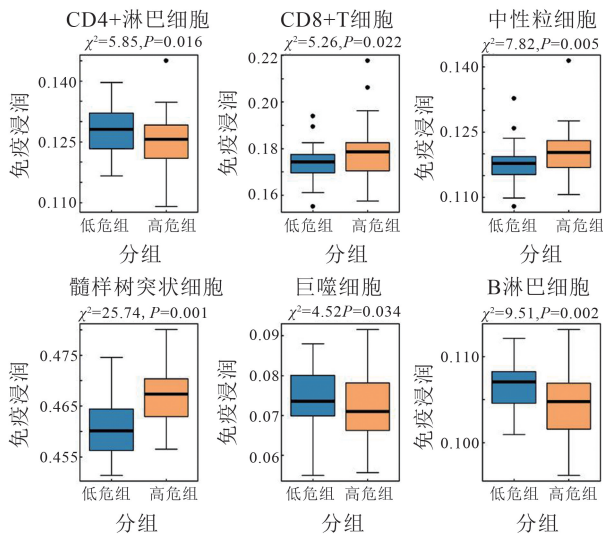


图 5 免疫浸润分析箱式图

者队列的鉴定和缺氧诱导疗法的实施<sup>[24]</sup>。本次研究筛选出 6 种浸润水平具有统计学差异的免疫细胞,其中中性粒细胞具有杀死胶质瘤细胞以及协调抗肿瘤的作用<sup>[25]</sup>,同时以中性粒细胞为介导的药物已经证实可以减缓胶质瘤的复发性生长<sup>[26]</sup>;巨噬细胞被证实会通过分泌 SPP1 来维持胶质瘤细胞的存活并刺激胶质瘤血管生成<sup>[27]</sup>;目前暂无发现髓样树突状细胞对人类胶质瘤进展影响研究,但 Dey 曾在实验中通过调节胶质瘤小鼠的髓样树突状细胞水平<sup>[28]</sup>,有效改善其生存状态并延长了生存时间,其免疫浸润丰度对人类胶质瘤的进展影响还有待进一步探索。

肿瘤的分子图谱往往蕴含着众多影响肿瘤进展的因素,包括细胞谱系、肿瘤微环境、细胞分化水平以及其他临床病理特征等,其中一些因素与疾病预后水平存在一定联系。survClust 相比传统无监督分型方法利用了这一关联,在深度挖掘预后相关的分子特征方面有着独特优势,更易获取具有显著生存差异的分子分型。在实际运用中该模型可用于患者的单个组学或多个组学数据整合的分型研究,要求不同患者间的组学数据类型匹配,此外也存在一定局限,如交叉验证框架下的多组学数据整合分型提高了计算成本,以及对不同组学的距离矩阵直接求平均,未能考虑不同组学间的异质性等,这些将作为今后算法改进的方向,同时可以考虑在生存结局加权的基础上,联合多种生物先验信息权重,进一步提高分型的准确性。

综上,本文基于 survClust 方法对中国人群胶质瘤多组学数据的整合分析,可以有效识别胶质瘤亚型,为胶质瘤的分型研究提供了新的实践依据,同时依据分型结果筛选得到的潜在生物标志物也为胶质瘤的个体化诊断和靶向药物研发提供了重要理论依据,后续需进一步结合分子生物学实验来验证其生物学意义。

参 考 文 献

[ 1 ] Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries[J]. CA Cancer J Clin, 2021, 71: 209-249.

[ 2 ] Louis DN, Perry A, Wesseling P, et al. The 2021 WHO Classification of Tumors of the Central Nervous System: a summary[J]. Neuro Oncol, 2021, 23: 1231-1251.

[ 3 ] Chen F, Wendl MC, Wyczalkowski MA, et al. Moving pan-cancer studies from basic research toward the clinic[J]. Nat Cancer, 2021, 2: 879-890.

[ 4 ] Xia M, Chen H, Chen T, et al. Transcriptional Networks Identify BRPF1 as a Potential Drug Target Based on Inflammatory Signature in Primary Lower-Grade Gliomas [ J ]. Front Oncol, 2021, 11: 766656.

[ 5 ] Giordano TJ. The Cancer Genome Atlas Research Network: A Sight to Behold[J]. Endocr Pathol, 2014, 25(4): 362-365.

[ 6 ] Yugi K, Kubota H, Hatano A, et al. Trans-Omics: How To Reconstruct Biochemical Networks Across Multiple “Omic” Layers [ J ]. Trends in Biotechnology, 2016, 34(4): 276-290.

[ 7 ] Zhao Z, Zhang KN, Wang Q, et al. Chinese Glioma Genome Atlas (CGGA): A Comprehensive Resource with Functional Genomic Data from Chinese Glioma Patients[J]. Genomics, Proteomics & Bioinformatics, 2021, 19(1): 1-12.

[ 8 ] Lock EF, Hoadley KA, Marron JS, et al. Joint and individual variation explained (JIVE) for integrated analysis of multiple data types [J]. The Annals of Applied Statistics, 2013, 7(1): 523-542.

[ 9 ] 李灵梅,魏亿芳,李治,等.基于 rMKL-LPP 方法的乳头状肾细胞癌多组学数据整合分型分析[J]. 中国卫生统计, 2022, 39(4): 522-528.

[ 10 ] Rappoport N, Shamir R. Multi-omic and multi-view clustering algorithms; review and cancer benchmark[J]. Nucleic Acids Research, 2018, 46(20): 10546-10562.

[ 11 ] Coretto P, Serra A, Tagliaferri R. Robust clustering of noisy high-dimensional gene expression data for patients subtyping[J]. Bioinformatics, 2018, 34(23): 4064-4072.

[ 12 ] Arora A, Olshen AB, Seshan VE, et al. Pan-cancer identification of clinically relevant genomic subtypes using outcome-weighted integrative clustering[J]. Genome Medicine, 2020, 12(1): 1-13.

[ 13 ] Kruskal J. Nonmetric multidimensional scaling: A numerical method [J]. Psychometrika, 1964, 29(2): 115-129.

[ 14 ] MacQueen, J. Some Methods for Classification and Analysis of Multivariate Observations [ J ]. Mathematical Statistics and Probability. 1967, 1: 281-297.

[ 15 ] Robert T, Guenther W, Trevor H. Estimating the Number of Clusters in a Data Set Via the Gap Statistic[J]. Journal of the Royal Statistical Society Series B: Statistical Methodology, 2001, 63(2): 411-423.

[ 16 ] Zhang Y, Yang Y. Cross-validation for selecting a model selection procedure[J]. Journal of Econometrics, 2015, 187(1): 95-112.

[ 17 ] Dweep H, Gretz N, Sticht C. MiRWalk database for miRNA-target interactions[J]. Methods Mol Biol, 2014, 1182: 289-305.

[ 18 ] Yu G, Wang LG, Han Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters[J]. Omics, 2012, 16(5): 284-287.

- [19] Ashburner M, Ball CA, Blake JA, et al. Gene ontology: Tool for the unification of biology[J]. Nat Genet,2000,25(1): 25-29.
- [20] Li T, Fu J, Zeng Z, et al. TIMER2.0 for analysis of tumor-infiltrating immune cells[J]. Nucleic Acids Res,2020, 48(W1): W509-W514.
- [21] Zhu C, Kros JM, Cheng C, et al. The contribution of tumor-associated macrophages in glioma neo-angiogenesis and implications for anti-angiogenic strategies[J]. Neuro Oncol,2017, 19(11): 1435-1446.
- [22] Savikko J, Rintala JM, Rintala S, et al. Epidermal growth factor receptor inhibition by erlotinib prevents vascular smooth muscle cell and monocyte-macrophage function in vitro[J]. Transpl Immunol, 2015,32(3): 175-178.
- [23] Edris B, Willingham SB, Weiskopf K, et al. Anti-KIT monoclonal antibody inhibits imatinib-resistant gastrointestinal stromal tumor growth[J]. Proc Natl Acad Sci U S A,2013,110(9): 3501-3506.
- [24] Thielpold AL, Lorenz NI, Foltyn M, et al. Mammalian target of rapamycin complex 1 activation sensitizes human glioma cells to hypoxia-induced cell death[J]. Brain,2017,140(10): 2623-2638.
- [25] Massara M, Persico P, Bonavita O, et al. Neutrophils in Gliomas[J]. Front Immunol,2017, 8: 1349.
- [26] Xue J, Zhao Z, Zhang L, et al. Neutrophil-mediated anticancer drug delivery for suppression of postoperative malignant glioma recurrence[J]. Nat Nanotechnol,2017,12(7): 692-700.
- [27] Chen P, Zhao D, Li J, et al. Symbiotic Macrophage-Glioma Cell Interactions Reveal Synthetic Lethality in PTEN-Null Glioma[J]. Cancer Cell,2019,35(6):868-884.
- [28] Dey M, Chang AL, Miska J, et al. Dendritic Cell-Based Vaccines that Utilize Myeloid Rather than Plasmacytoid Cells Offer a Superior Survival Advantage in Malignant Glioma[J]. J Immunol,2015, 195(1): 367-376.

(责任编辑:郭海强)

(上接第 643 页)

在 20 世纪 80 年代,我们出现了第一批博士点,包括上海医科大学、第四军医大学、华西医科大学、中山医科大学。他们的学科带头人许世瑾、郭祖超、杨树勤和胡孟璇,我们不能忘记。许世瑾、郭祖超已经介绍了。刚才有人讲了,胡孟璇教授出国回来,自己花了几千美金买了一台计算机,给教研室大家公用,是我们中山医科大学第一台。杨树勤是我们《卫生统计学》教材第一版到第三版的主编。现如今我们梯队建设后继有人,我们的硕士点不稀奇,极普遍,在座的很多学校都有博士点。我们的教师都有高学位和海外留学经历,我们各校都有一支强劲的队伍,国外的医学院校无可比拟。

第二个传统是“合作成风,劲往一处使”。今天就是个佐证,七个学会一起来开年会。我们《卫生统计学》教材是一个很典型的产物,现在已经到第九版了。以我主编的第五、六、七版为例,作者有 30 位,来自 27 所医学院校,这就是合作,我们集体编,大家用。

我们的研究生教材和手册在国外出版,最早是我和苏炳华、徐勇勇合编第一版《医学统计与电脑实验》,现在我们又出了第二本和第三本:《生物医学研究中的统计方法》和《医学统计学手册》。第一本是我写信要求国外出版社出的,信中介绍了我们中国的教

学经验;而第二本、第三本则是国外出版社来信、来人邀请我主编,因为全世界都知道中国的医学统计学力量强大。

第三个传统:“东西南北中,同行一家亲”。在我们行内,有任何机遇都是彼此眷顾的。以自然科学基金评审为例,我们的年轻学者,不管你是谁的学生,不管你是哪个学校的,我们所有的专家看到了申请书里边有新的东西,前景很好,我们就马上给予扶持。还有,我们一方有需要,大家都去帮。山东大学办大数据与数据科学专业,前不久,我亲自飞过去了,陈峰也去了,王彤和郝元涛没有空,在线上发言,帮他们出主意。因为这是一个好的苗子,好的前景,我们是以学科为重,互相扶持的。

我们七个学会在一起开会,坐在一起筹备会议,我们有共同的品质:virtuous and sincere。请记住这两个英文单词,怎么翻译?各位想一想,我的翻译最确切,两个字:厚道!

最后,我们小结一下,要“三学习”:学数学、现代计算技术、服务对象;“三传统”:尊师爱徒、合作共赢、同行厚道。

(文字整理:郝元涛、顾菁、朱淑明)

(责任编辑:郭海强)