

网络出版时间: 2021-4-6 9:06 网络出版地址: <https://kns.cnki.net/kcms/detail/34.1065.r.20210402.1340.017.html>

基于 TCGA 数据库构建肾透明细胞癌自噬相关基因风险预测模型

陈俊逸, 陈晶, 张蒙, 梁朝朝, 陈先国

摘要 目的 基于 TCGA 数据库中差异表达自噬相关基因 (DEARGs) 构建肾透明细胞癌 (ccRCC) 预后预测模型。方法 从 TCGA 数据库中下载 537 例 ccRCC 及 96 例正常组织 RNA 测序数据, 提取自噬相关基因 (ARGs) 表达谱。通过对比癌和正常组织, 获取差异表达基因列表, 将其中 489 例有配对临床信息的 ccRCC 样本通过随机抽样的方法按照 7:3 的比例拆分为训练组 ($n=344$) 与验证组 ($n=145$), 在训练组中利用单因素和多因素 Cox 回归分析构建预后预测模型, 并在验证组中对模型进行验证评估。结果 分析筛选出 36 个 DEARGs, 运用单因素和多因素 Cox 回归分析筛选得 8 个 DEARGs (CX3CLI、PRKCQ、RAB、VMP1、IFNG、EIF4EBP1、ATG16L2 和 BID) 及其风险系数 (coefficient, COEF), 并计算每位患者的风险评分 (RS), 以中位风险评分 (RS) 为截断值, 将训练组和验证组的患者分为低风险组和高风险组。Kaplan-Meier 法和 Log-Rank 法分析显示, RS 高的患者总体生存率差于评分低的患者 ($P<0.05$)。研究在验证组得到一致性的结果 ($P<0.05$)。此外, 多因素回归分析揭示该预测模型是 ccRCC 患者预后预测的独立风险因子 ($P<0.05$)。结论 自噬相关差异表达基因是 ccRCC 患者潜在的诊断和预后评估标志物, 可为患者术后个体化治疗提供理论支持。关键词 肾透明细胞癌; 自噬相关基因; TCGA 数据库; 预后中图分类号 R 737.11

文献标志码 A 文章编号 1000-1492(2021)05-0757-05
doi: 10.19405/j.cnki.issn1000-1492.2021.05.016

肾透明细胞癌 (clear cell renal cell carcinoma, ccRCC) 是泌尿系统最常见的恶性肿瘤之一^[1], 其发病率占肾细胞癌 (renal cell carcinoma, RCC) 发病率的 70%~80%。手术是局限性 RCC 的主要治疗方式, 而晚期和转移性 RCC 主要依赖于免疫治疗、靶向药物和化疗^[2]。该疾病的病死率仍呈逐年上升趋势^[3]。因此, 鉴定出新型的生物标志物可用于 ccRCC 患者的复发或死亡风险预测, 对于患者的术

后治疗意义重大。

自噬是一种重要的生物学现象, 在自噬相关基因 (autophagy-related genes, ARG) 的调控下, 细胞可通过分解大分子物质^[4] 提供氨基酸及合成新蛋白的其他必需因子。已有大量研究^[5] 支持自噬与 ccRCC 的发生和进展密切相关。然而, 关于 ARGs 是否可用于 ccRCC 患者术后预后评估尚未明确。

1 材料与方法

1.1 数据资料 从 TCGA 中下载 ccRCC 患者的 ARGs 表达谱数据及临床数据, 包括 537 例 ccRCC 肿瘤和 96 例非肿瘤组织的 RNA 测序数据, 其中包括 489 例有配对临床资料的 ccRCC 样本。提取的临床数据信息包括总生存期 (overall survival, OS)、年龄、性别、肿瘤分级、T 分期、病理分级。所有数据下载于 2020 年 4 月 12 日。

1.2 方法 应用 R 统计软件中的 “EdgeR” 软件包对下载的数据进行提取分析, 筛选出 ccRCC 和非肿瘤组织间的差异表达自噬相关基因 (differentially expressed autophagy-related genes, DEARGs) (筛选条件为 $\log_2 \text{FoldChange} \geq 2$ 或 ≤ -2 , $\text{FDR } P < 0.05$)。然后, 对 DEARGs 采用单因素 Cox 回归分析, 依据 $P < 0.05$ 筛选出与患者 OS 相关的 DEARGs。接下来, 将筛选出的与预后相关的 DEARGs 纳入多因素 Cox 回归分析中并建立预后风险评分 RS (risk score, RS) 模型, 计算每位患者的 RS, 依据评分中位值将患者分为高、低风险组, 利用卡普兰-梅尔估计量 (Kaplan-Meier, K-M) 法对比组间预后的差异, 并计算受试者工作特征 (receiver operating characteristic, ROC) 曲线评价模型的有效性, 并计算曲线下面积 (area under curve, AUC)。最后通过多因素 Cox 回归分析探讨该 RS 是否独立于各项临床病理参数 (年龄、性别、肿瘤分级、T 分期、病理分级), 作为 ccRCC 患者独立预后因子。

2 结果

2.1 筛选 DEARGs 从 TCGA 数据库中获取了 537

2020-12-11 接收

基金项目: 国家自然科学基金 (编号: 81970597)

作者单位: 安徽医科大学第一附属医院泌尿外科, 合肥 230022

作者简介: 陈俊逸, 男, 硕士研究生;

陈先国, 男, 副教授, 硕士生导师, 责任作者, E-mail:

cxg7866186@126.com

例 ccRCC 肿瘤和 96 例非肿瘤组织的 RNA 测序数据,提取了 215 个 ARGs 的表达谱。利用“EdgeR”软件对比了癌和非癌组织,以 $\log\text{FoldChange} \geq 2$ 或 ≤ -2 和 $\text{FDR } P < 0.05$ 为截断标准,共筛选出了 36 个上调和 9 个下调的 ARGs。图 1A 散点图显示了 9 个下调基因(*FAM215A*、*DIRAS3*、*PRKCQ*、*GABARA-PL1*、*ERBB2*、*BAG1*、*HIF1A*、*TP63* 和 *MTOR*) 和 36 个上调基因(*CX3CLI*、*ATG12*、*BID*、*IL24*、*RACK1*、*FAS*、*BAX*、*CASP4*、*VMP1*、*CCR2*、*P4HB*、*GAPDH*、*ERO1A*、*GRID1*、*EGFR*、*MYC*、*BNIP3*、*SERPINA1*、*SPHK1*、*RAB24*、*RGS19*、*CASP1*、*NLRG4*、*NRG3*、*APOL1*、*EIF4EBP1*、*HSPB8*、*ATG16L2*、*BIRC5*、*CXCR4*、*ATG9B*、*TP73*、*NKX23*、*VEGFA*、*IFNG* 和 *CDKN2A*)。图 1B 热图显示 DEARGs 表达水平的层次聚类。

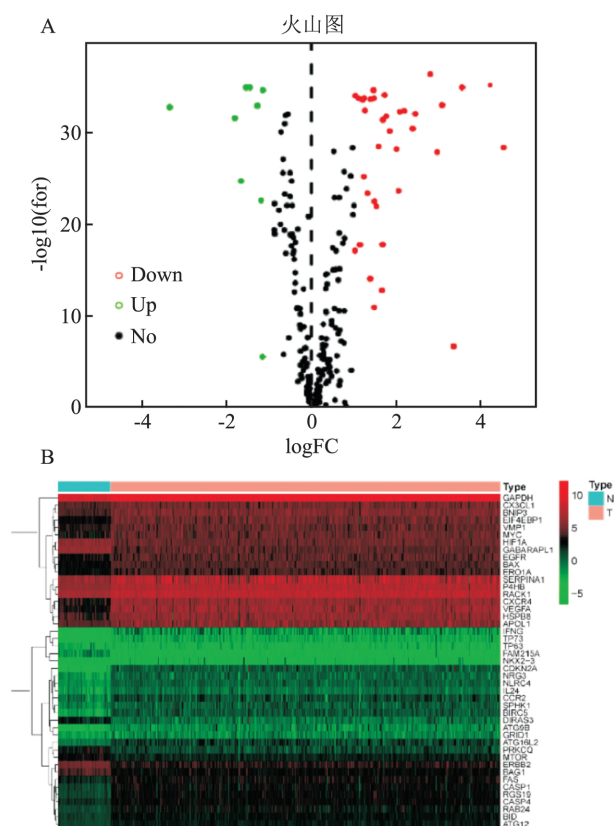


图 1 ccRCC 组织与正常组织之间 DEARGs

A: 火山图揭示 ccRCC 中 215 个 ARGs 的表达水平; B: 热图显示 DEARGs 表达水平的层次聚类; 红色: 高表达; 绿色: 低表达; 黑色: 基因在 2 组织之间没有差异

2.2 基于 DEARGs 的预后模型的建立 将有配对临床资料的 489 例 ccRCC 样本(表 1),通过随机抽样的方法按照 7:3 的拆分为训练组($n = 344$)与验证组($n = 145$),训练组用于建模,验证组用于外部验证预后模型的效能。通过对这 45 个 DEARGs 的

单因素 Cox 回归分析后,得到了 21 个与 OS 有关的 DEARGs(图 2),纳入多因素 Cox 回归分析,筛选出 8 个 DEARGs 用于模型的构建(*BID*、*EIF4EBP1*、*ATG16L2*、*CX3CLI*、*IFNG*、*PRKCQ*、*EIF4EBP1*、*VMP1*)。其中, *ATG16L2*、*RAB24*、*EIF4EBP1*、*IFNG* 和 *BID* 为高风险基因,而 *VMP1*、*PRKCQ* 和 *CX3CLI* 为保护基因。各基因的风险系数(coefficient, COEF)及风险比(hazard ratio, HR)见表 2。COEF $\times$ 风险基因表达值为 RS 的因变量,得到 OS RS = (0.718 $\times$ *BID* 表达值) + (0.281 $\times$ *EIF4EBP1* 表达值) + (0.326 $\times$ *ATG16L2* 的表达值) + (-0.220 $\times$ *CX3CLI* 表达值) + (0.280 $\times$ *IFNG* 表达值) + (-0.299 $\times$ *PRKCQ* 表达值) + (0.281 $\times$ *EIF4EBP1* 表达值) + (-0.233 $\times$ *VMP1* 表达值)。

2.3 预测模型的有效性和稳定性评估 训练组中位 RS 为 0.95,将评分 < 0.95 的 172 例患者纳入低风险组,评分 ≥ 0.95 的 172 例患者纳入高风险组。训练组中 ccRCC 患者的 RS 评分、生存分布和 ARGs 表达见图 3,本研究发现随着 RS 值的升高,患者生

表 1 TCGA 数据库中 489 例 ccRCC 患者的临床特征

临床特征	n(%)
年龄(岁)	
≤ 60	230(47)
> 60	259(53)
性别	
男	
女	
肿瘤分级	
G1	14(3)
G2	230(47)
G3	207(42)
G4	38(8)
T 分期	
T1	273(56)
T2	68(14)
T3	137(28)
T4	11(2)
病理分级	
Stage III	122(25)
Stage IV	367(75)

表 2 DEARGs 多因素 Cox 回归分析

基因	coef	HR	P 值
<i>BID</i>	0.718	2.262 90	0.002 475
<i>CX3CLI</i>	-0.220	0.645 71	0.071 083
<i>IFNG</i>	0.280	1.379 18	0.050 049
<i>PRKCQ</i>	-0.433	0.714 52	0.004 970
<i>EIF4EBP1</i>	0.281	1.594 57	0.004 193
<i>VMP1</i>	-0.233	0.680 42	0.020 351
<i>ATG16L2</i>	0.326	1.449 69	0.002 136
<i>RAB24</i>	0.299	1.385 77	0.002 423

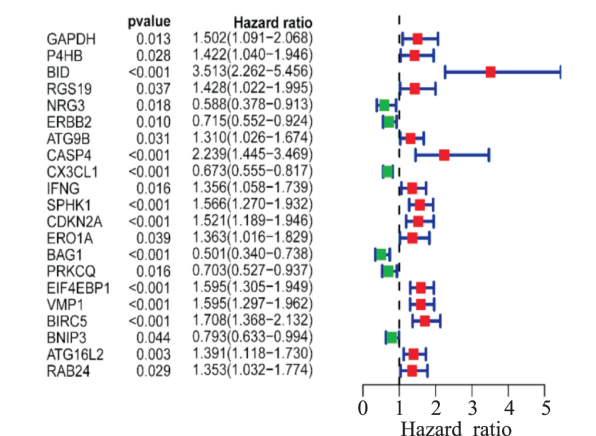


图2 单因素 Cox 回归中 21 个与总体生存有关的 ARGs 森林图

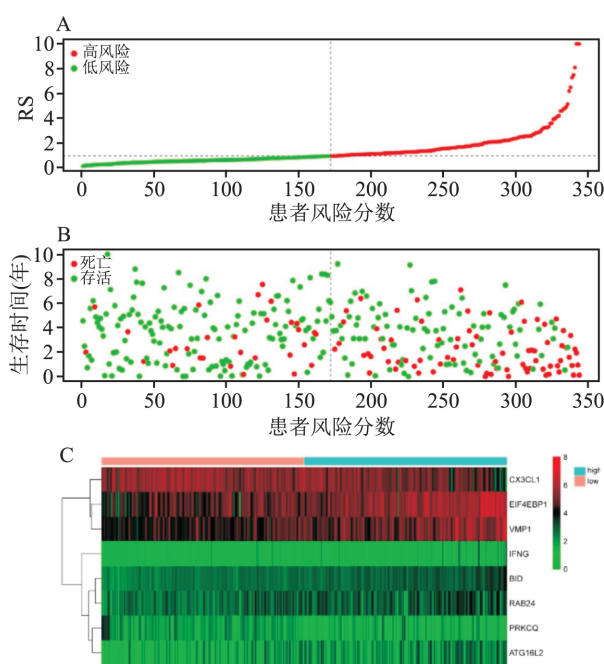


图3 RS 分布图与 DEARGs 热图及生存时间散点图

A: 高风险(红色)和低风险(绿色) ccRCC 患者的 RS 分布; B: 散点图显示了预后模型中 ccRCC 患者的生存状况; 红点: 死亡的患者; 绿点: 生存的患者; C: 高风险(蓝色)和低风险(粉色) ccRCC 患者中风险基因的表达; 绿色: 最低表达; 红色: 最高表达

存时间呈缩短趋势,且死亡患者人数明显增加(图中红点显示)。热图中高 RS 患者显示 *VMP1*、*ATG16L2*、*RAB24*、*BID* 和 *EIF4EBP1* 上调,而 *IFNG*、*CX3CL1* 和 *PRKCQ* 降低,低风险患者显示 *IFNG*、*CX3CL1* 和 *PRKCQ* 上调,*VMP1*、*ATG16L2*、*RAB24*、*BID* 和 *EIF4EBP1* 上调。进一步 K-M 法结果显示,高风险组患者的预后比低风险组差(图 4A, $P < 0.05$),且 ROC 曲线提示该模型具有较高的预测价值(图 4B, $AUC = 0.783$, $95\% CI = 0.723 \sim 0.897$)。为验证预后预测模型的灵敏度与稳定性,对验证组

进行了分析。依据 RS 评分,将验证组 145 例 ccRCC 患者分为高风险($n = 72$)和低风险($n = 73$)组, K-M 法揭示低风险组患者与高风险组相比 OS 延长($P < 0.05$)(图 5)。ROC 曲线提示该模型在验证队列中仍持有较高的预测价值($AUC = 0.754$, $95\% CI = 0.706 \sim 0.873$),特异度为 84.1%。综上所述,这些结果提示 DEARGs 预测模型能够有效的预测 ccRCC 患者的预后。

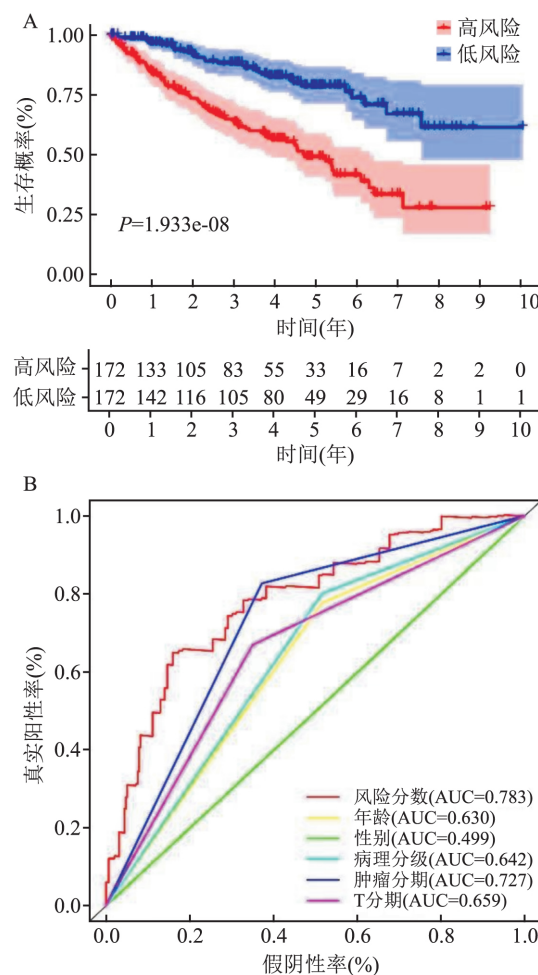


图4 训练组 DEARGs 模型预测效果评估

A: 训练组高风险组(红线)和低风险组(蓝线) K-M 法生存曲线分析; B: 训练组各临床指标与 RS 的 ROC 曲线分析比较

2.4 自噬相关预后模型与患者生存期独立相关研究进一步利用多因素分析评估了训练组 RS 与临床病理参数(年龄、性别、肿瘤分期、T 分期、病理分级)之间的关系。如表 3 所示,单因素 Cox 回归分析表明 ccRCC 患者的年龄、性别、肿瘤分期、病理分级、RS 与 OS 相关($P < 0.05$)。多因素 Cox 回归分析揭示 RS 与总体生存相关,可作为患者 OS 评估的独立预测因子($P < 0.05$)。此外,病理分级也与患

者 OS 相关 ($P < 0.05$), 可作为预后评估独立因子。这些结果表明, 基于 DEARGs 的 OS 预后预测模型可有效区分高危和低危死亡风险的 ccRCC 患者, 且可作为预后评估的独立预测因子。

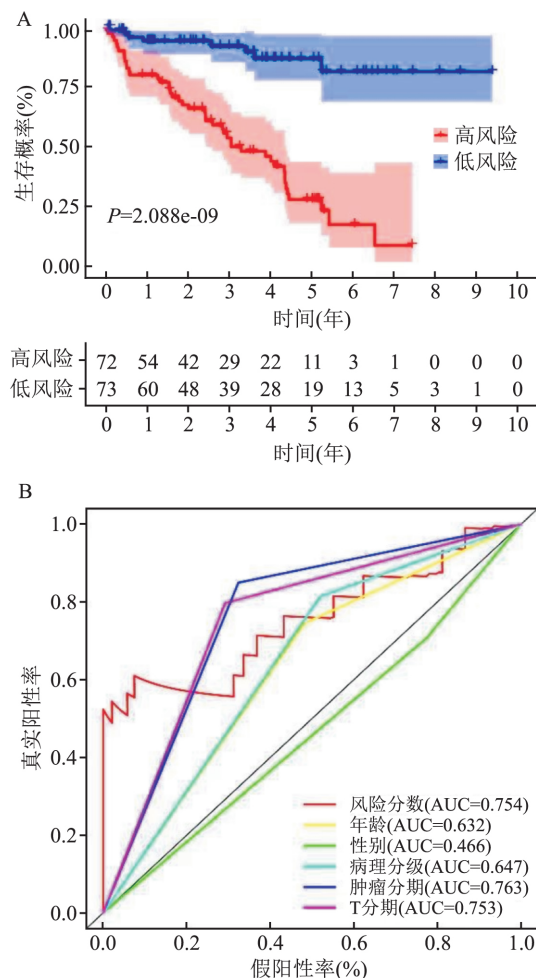


图5 验证组 DEARGs 模型预测效果评估

A: 验证组高风险组(红线)和低风险组(蓝线) ccRCC 患者 K-M 法生存曲线分析; B: 验证组各临床指标与 RS 的 ROC 曲线分析比较

表3 单因素和多因素 Cox 回归分析影响患者总体生存期的因素

临床指标	单因素回归分析		多因素回归分析	
	HR(95% CI)	P 值	HR(95% CI)	P 值
年龄(岁)	2.089 288(1.39~3.14)	0.000 381	1.799 759(1.18~2.75)	0.006 625
性别	0.791 383(0.54~1.16)	0.227 669	0.970 347(0.66~1.43)	0.879 874
肿瘤分级	2.257 457(1.49~3.43)	0.000 133	1.485 576(0.95~2.32)	0.081 922
病理分级	3.449 428(2.31~5.14)	1.21E-09	5.845 534(2.78~12.30)	3.26E-06
T分期	2.710 965(1.85~3.98)	3.41E-07	0.385 932(0.19~0.79)	0.009 365
RS	1.207 111(1.15~1.27)	3.67E-44	1.158 346(1.10~1.21)	5.15E-09

3 讨论

有研究^[8]表明了多个 ARGs 与 ccRCC 发病或进展之间的关联, 但尚无系统研究评估 ARGs 是否

可以作为 ccRCC 患者预后评估标志物。本研究从 TCGA 数据库中 ccRCC 和正常组织表达谱与临床生存资料进行统计分析得到基于 8 个 AGEs 的多因素 RS 模型: $RS = (0.718 \times BID \text{ 表达值}) + (0.281 \times EIF4EBP1 \text{ 表达值}) + (0.326 \times ATG16L2 \text{ 的表达值}) + (-0.220 \times CX3CLI \text{ 表达值}) + (0.280 \times IFNG \text{ 表达值}) + (-0.299 \times PRKCQ \text{ 表达值}) + (0.281 \times EIF4EBP1 \text{ 表达值}) + (-0.233 \times VMP1 \text{ 表达值})$ 。K-M 法、log-rank 法和 ROC 分析在训练组和验证组中揭示了预测模型的特异性和稳定性。同时, 多因素分析揭示该模型为 ccRCC 患者 OS 的独立预测因素, 进一步佐证了 ARGs 与 ccRCC 的关联。

模型共纳入 8 个差异表达的 ARGs。BID 位于染色体 22q11.21 上, 编码与凋亡相关的蛋白, 研究^[9]显示其可在甲状腺癌组织中高表达, 并与患者预后紧密相关。ATG16L2 位于 11 号染色体上, 尚无文献报道该基因在肿瘤发生或进展中的作用。而在当前研究中, 显示 ATG16L2 表达上调并与 ccRCC 患者的 OS 降低有关。EIF4EBP1 在肝细胞癌和乳腺癌组织中表达水平升高且与患者预后不佳有关^[10]。PRKCQ 在肿瘤中的作用机制尚未完全阐明, 研究^[11]表明, PRKCQ 在包括胃肠道间质瘤在内的实体肿瘤中均表达上调。RAB24 目前在功能方面研究较少, Chen et al^[12]发现其在肝细胞癌中上调, 过表达该基因可促进肝癌细胞进展和上皮-间质转化过程。CX3CLI 基因位于 16 号染色体, 研究^[13]表明 CX3CLI 与 CCR1, CCL5 受体协同诱导结直肠癌的肝转移。VMP1 在肝细胞癌中, 抑制增殖和转移, 而在胰腺细胞系中, 对 VMP1 的抑制会降低细胞凋亡^[14]。IFNG 在肿瘤方面研究较少, Garcia-tunon et al^[15]发现 IFNG 在乳腺纤维囊性病变、原位和浸润性肿瘤中的表达量上调, 提出 IFNG 可作为乳腺癌的一种前瞻性治疗方式。

综上所述, 本研究通过 Cox 回归分析组建了由 8 个 DEARGs 构成的 ccRCC 患者预后预测模型。研究证明了该模型具有较高的特异性与灵敏度, 其可补充临床病理参数用于术后患者的预后评估和治疗指导, 推进了 ccRCC 患者的个体化医疗进展。

参考文献

- Miller K D, Nogueira L, Mariotto A B, et al. Cancer treatment and survivorship statistics [J]. CA Cancer J Clin, 2019, 69(5): 363-85.
- Chen V J, Hernandez-Meza G, Agrawal P, et al. Time on therapy for at least three months correlates with overall survival in metastat-

- ic renal cell carcinoma [1]. *Cancers* (Basel), 2019, 11(7): 1000.
- [5] Siegel R L, Miller K D, Jemal A. Cancer statistics, 2019 [J]. *CA Cancer J Clin*, 2019, 69(1): 7–34.
- [6] Dikic I, Johansen T, Kirkin V. Selective autophagy in cancer development and therapy [J]. *Cancer Res*, 2010, 70(9): 3431–4.
- [7] Cost N G, Czyzyk-Krzeska M F. Regulation of autophagy by two products of one gene: TRPM3 and miR-204 [J]. *Mol Cell Oncol*, 2015, 2(4): 100207.
- [8] Lin P, He Y, Wen D Y, et al. Comprehensive analysis of the clinical significance and prospective molecular mechanisms of differentially expressed autophagy-related genes in thyroid cancer [J]. *Int J Oncol*, 2018, 53(2): 603–19.
- [9] Rutkovsky A C, Yeh E S, Guest S T, et al. Eukaryotic initiation factor 4E-binding protein as an oncogene in breast cancer [J]. *BMC Cancer*, 2019, 19(1): 491.
- [10] Blay P, Astudillo A, Buesa J M, et al. Protein kinase C theta is highly expressed in gastrointestinal stromal tumors but not in other mesenchymal neoplasias [J]. *Clin Cancer Res*, 2004, 10(12 Pt 1): 4089–95.
- [11] Chen Z, Wang X, Liu R, et al. KDM4B-mediated epigenetic silencing of miRNA-615-5p augments RAB24 to facilitate malignancy of hepatoma cells [J]. *Oncotarget*, 2017, 8(11): 17712–25.
- [12] Isik A, Okan I, Firat D, et al. A new prognostic strategy for gastric carcinoma: albumin level and metastatic lymph node ratio [J]. *Minerva Chir*, 2014, 69(3): 147–53.
- [13] Pagliarini V, Wirawan E, Romagnoli A, et al. Proteolysis of ambra1 during apoptosis has a role in the inhibition of the autophagic pro-survival response [J]. *Cell Death Differ*, 2012, 19(9): 1495–504.
- [14] Garcia-tunon I, Ricote M, Ruiz A A, et al. Influence of IFN-gamma and its receptors in human breast cancer [J]. *BMC Cancer*, 2007, 7(1): 158.

Construction of a risk prediction model for autophagy-related gene prediction model for renal clear cell carcinoma based on TCGA database

Chen Junyi, Chen Jing, Zhang Meng, et al

(Dept of Urology, The First Affiliated Hospital of Anhui Medical University, Hefei 230022)

Abstract Objective To establish a prognostic model for renal clear cell carcinoma based on differentially expressed autophagy related genes in TCGA database. **Methods** The RNA sequencing data for 537 cases of renal clear cell carcinoma and 72 cases of normal tissues were downloaded from the TCGA database for extracting the expression profiles of autophagy related genes. A comparison between carcinoma and normal tissue was conducted to obtain a list of differentially expressed genes, in which 489 samples of renal clear cell carcinoma with paired clinical information were randomized 7 : 3 to training group($n = 344$) and validation group($n = 145$). For the training group, the univariate and multivariate Cox regression analyses were used to construct the prognostic model, which was then validated by the validation group. **Results** Total 36 differentially expressed autophagy-related genes were screened out. The univariate and multivariate Cox regression analyses screened out 8 differentially expressed autophagy-related genes(*CX3CLI*, *PRKCQ*, *RAB*, *VMP1*, *IFNG*, *EIF4EBP1*, *ATG16L2* and *BID*) and relevant risk factors, and the risk score was calculated for each patient. With the median risk score as the cutoff value, the patients in the training group and validation group were divided into low risk group and high risk group. The Kaplan-Meier and log-rank analyses showed that the overall survival rate of patients with high risk scores was lower than that of patients with low risk scores($P < 0.05$). The results obtained in the validation group were consistent($P < 0.05$). In addition, the multivariate regression analysis revealed that the prognostic model was an independent risk factor for predicting the prognosis of renal clear cell carcinoma($P < 0.05$). **Conclusion** The differentially expressed autophagy-related genes are potential diagnostic and prognostic markers in patients with renal clear cell carcinoma, which may support individualized postoperative treatment theoretically.

Key words clear cell renal cell carcinoma; autophagy-related differentially expressed gene; TCGA database; prognosis