

# 胃癌顺铂抵抗相关免疫标记基因的富集分析

汪圣毅 程彦 李旭升 闫亚飞 张尚鑫 闫强 李永翔

**摘要** 目的 采用基因集富集分析(GSEA)方法寻找胃癌抵抗顺铂的免疫标记基因(ISGs)。方法 用GEO数据库的GSE94714数据集,GEO2R分析差异基因,观察条件筛选对基因数的作用,GSEA纳入耐药、未耐药组胃癌细胞的全部差异表达基因,与分子标签数据库比较,获取ISGs,交集筛选,Kaplan Meier Plotter分析交集基因对胃癌预后的影响。结果 差异表达基因共34 183个,其中上调12 452个、下调17 381个,筛选差异倍数的增加使排除基因数增加。GSEA富集到标准化富集评分(NES)排序前6的条目( $P < 0.01$ ),其中的交集基因包括线粒体核糖体蛋白L12、富含脯氨酸蛋白13、毛状蛋白样F-肌动蛋白结合蛋白1、聚(RC)结合蛋白1、艾杜糖-2-硫酸酯酶、LIM结构域2、富含嘌呤元素结合蛋白A、小视觉叶同源物、CCR4-非转录复合物亚单位3、转化生长因子 $\beta$ 1、二酰甘油激酶 $\zeta$ 、接头蛋白2,12个基因与胃癌的总生存时间有关,均具有统计学意义( $P < 0.05$ )。结论 GSEA方法可有效获取胃癌顺铂抵抗的ISGs,新发现的基因作为潜在靶点,可促进胃癌化疗抵抗机制的研究。

**关键词** 胃癌; 化疗抵抗; 基因集富集分析; 基因表达数据库; 免疫标记基因; 预后

中图分类号 R 735.2

文献标志码 A 文章编号 1000-1492(2020)02-0200-05  
doi: 10.19405/j.cnki.issn1000-1492.2020.02.009

胃癌在我国高发,手术是主要治疗方法,辅助化疗可改善预后<sup>[1]</sup>,但化疗抵抗制约疗效,与基因、通路的作用有关<sup>[2-3]</sup>,免疫标记基因(immunologic signature genes, ISGs)与化疗抵抗有关<sup>[4]</sup>。外周髓磷脂蛋白22(peripheral myelin protein 22, PMP22)增强胃癌顺铂抵抗<sup>[3]</sup>,是否通过ISGs调控尚不清楚。常规富集分析可发现PMP22相关的化疗抵抗通路和基因<sup>[2]</sup>,但预设差异基因的倍数和统计 $P$ 值,不能较为全面地反映整体信息<sup>[5]</sup>。基因集富集分析(gene set enrichment analysis, GSEA)纳入全部的差

异基因,具有不预设组间差异基因的优点<sup>[6]</sup>,可发现细微变化。为明确ISGs与PMP22相关胃癌顺铂抵抗的关系,用GSEA法分析GSE 94714数据集,寻找PMP22下游的ISGs,验证新发现ISGs对胃癌总生存的影响,为深入揭示胃癌化疗抵抗的机制提供参考。

## 1 材料与方法

**1.1 表达谱数据** 用基因表达数据库(gene expression omnibus, GEO)的GSE 94714数据集,包含6个样本,3个样本(GSM2481329 ~ GSM2481331)的胃癌细胞MGC-803进行了PMP22基因敲减,另外3个样本(GSM2481332 ~ GSM 2481334)的PMP22未行敲减处理。GEO2R分析差异基因,设置分组: PMP22基因敲减3个样本为A组,未敲减3个样本为B组,保存所有差异基因至excel表,得到差异基因表。

**1.2 芯片数据及探针基因名转换** 下载GSE94714的soft文件,导入excel,复制探针和对应基因名称(GENE\_SYMBOL),形成新的excel表即芯片数据表,探针名称升序排序。差异基因表中的探针名称进行相同的升序排序,复制ID、adj. P. Val、logFC共3列到芯片数据表,excel数据高级筛选,匹配出探针对应的基因名称,包含差异基因表,形成探针基因名转换表,定位GENE\_SYMBOL列的空值,删除没有GENE\_SYMBOL的所有行,另存为包含GENE\_SYMBOL的差异基因表,用于后续分析。

**1.3 差异表达分析** 包含GENE\_SYMBOL的差异基因表,导入在线分析工具(Easy Tools v 2.2 by Chris Lou)中([www.chrislifescience.club:3838/R/AnnoE2](http://www.chrislifescience.club:3838/R/AnnoE2))的Volcano Online,绘制火山图。差异倍数(fold change, FC)以2为底的对数( $\log_2$  FC或Log FC)分别设置为1、1.5,  $P$ 值 $< 0.05$ ,计算两种Log FC条件下,有差异(上调、下调)和无差异的基因数目,并计算无差异基因数目的差值。

**1.4 GSEA** 未筛选的所有差异基因,用含GENE\_SYMBOL的差异基因表,导入在线工具([www.chrislifescience.club:3838/R/AnnoE2](http://www.chrislifescience.club:3838/R/AnnoE2))的GSEA中,列名

2019-11-22 接收

基金项目: 国家自然科学基金(编号: 81874063)

作者单位: 安徽医科大学第一附属医院普外科胃肠外科,合肥 230022

作者简介: 汪圣毅,男,副教授,主任医师,硕士生导师;

李永翔,男,教授,主任医师,博士生导师,责任作者, E-mail: 1874010558@qq.com

中分别填写导入表格中对应各列的名称,分子标签数据库(MSigDB, molecular signatures database)中ID选择c7。标准化富集评分(normalized enrichment score, NES)降序排列,取NES前6位的基因集进一步分析。

**1.5 免疫标记基因集的交集基因** NES前6位的ISGs条目导入MSigDB(<http://software.broadinstitute.org/gsea/msigdb/search.jsp>),查询并下载基因,在线工具分析交集基因(<http://bioinformatics.psb.ugent.be/cgi-bin/liste/Venn/>),绘制韦恩图。

**1.6 统计学处理** 在线工具的R软件进行统计,富集评分标准化的值(NES)进行排列检验(permutation test),错误发现率(false discovery rate, FDR)方法判断NES的统计学意义;组间基因表达差异进行t检验;交集基因导入KM-plotter数据库(<http://km-plot.com/analysis/>),根据基因表达水平自动选择临界值,分割为高、低表达组, Kaplan-Meier法分析基因表达水平与胃癌患者总生存(overall survival, OS)的关系, Log-rank检验比较组间差异,给出风险比(hazard ratio, HR)、95%的可信区间(confidence interval, CI),  $P < 0.05$ 为差异有统计学意义。

**2 结果**

**2.1 差异表达矩阵** GEO2R分析共获取差异基因34 183个,用芯片数据表对表达数据进行整理,转换基因名,删除无GENE\_SYMBOL的行,获差异表达基因共29 833个。COUNTIF函数计算Log FC > 0的上调基因共12 452个, Log FC < 0的下调基因共17 381个。

**2.2 差异基因表达** 差异倍数的对数(Log FC)为1时,结果表格中, COUNTIF函数计算上调(UP)、下调(DOWN)、无差异(NOT)基因数目,分别为3 819个、3 730个、22 284个; Log FC为1.5时, UP、DOWN、NOT基因数分别为2 300个、1 492个、26 041个。两种条件下的火山图见图1。两种条件下, UP、DOWN、NOT基因数的差值分别为1 519个、2 238

个、3 757个。上下调基因数目与筛选条件的关系见图2。

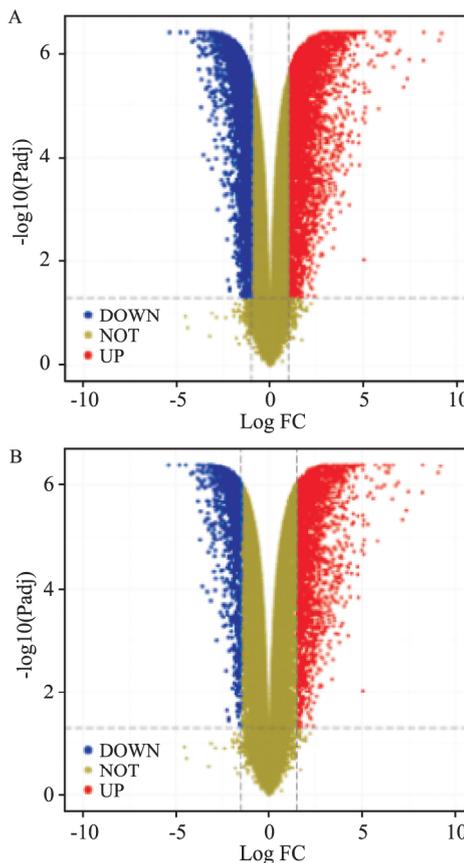


图1 差异表达基因的火山图

A: Log FC为1  $P < 0.05$ 的火山图; B: Log FC为1.5  $P < 0.05$ 的火山图

**2.3 富集分析** MSigDB中的c7作为富集分析的基因集,富集得到条目4 872个, NES降序排序,排序前6的结果见表1,前6位的条目作GSEA图,见图3。

**2.4 免疫标记基因集的交集基因** NES前6位的ISGs条目导入MSigDB,查询、下载基因,在线工具分析交集, A、C、D基因集得到的交集基因1个,为线粒体核糖体蛋白L12(mitochondrial ribosomal protein L12, MRPL12),见图4a。E、F基因集的交集基

表1 基因集富集分析NES前6的条目

条目编码	ID	NES	P值
A	GSE9006_HEALTHY_VS_TYPE_2_DIABETES_PPMC_AT_DX_UP	2.684	0.004
B	GSE9006_TYPE_1_VS_TYPE_2_DIABETES_PPMC_AT_DX_UP	2.626	0.004
C	GSE9006_TYPE_1_DIABETES_AT_DX_VS_1MONTH_POST_DX_PPMC_UP	2.532	0.004
D	GSE42724_NAIVE_VS_B1_BCELL_DN	2.426	0.004
E	GSE6269_FLU_VS_E_COLI_INF_PPMC_UP	2.399	0.003
F	GSE9988_LPS_VS_VEHICLE_TREATED_MONOCYTE_DN	2.368	0.004

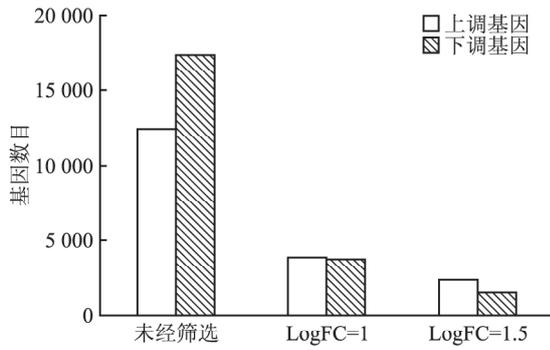


图2 上调和下调基因数与筛选条件的关系

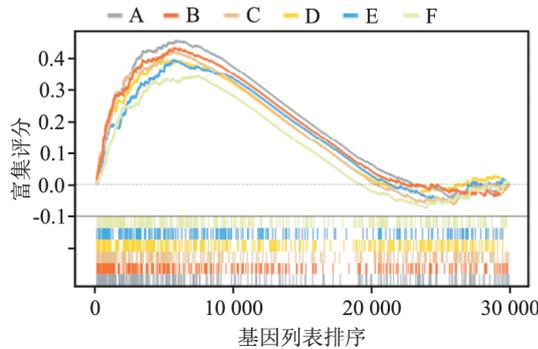


图3 前6位免疫标记相关条目的GSEA  
A~F: 对应表1中条目编码

因 2 个: 富含脯氨酸蛋白 13 ( proline rich 13 , PRR13)、富含嘌呤元素结合蛋白 A( purine rich element binding protein A ,PURA) ; A、E 的交集基因 5 个: 艾杜糖 2-硫酸酯酶( iduronate 2-sulfatase ,IDS)、LIM 结构域 2( LIM domain containing 2 ,LIMD2)、小视觉叶同源物( small optic lobes homolog ,SOLH)、CCR4-非转录复合物亚单位 3( CCR 4-NOT transcription complex subunit 3 ,CNOT3)、转化生长因子  $\beta$ 1 ( transforming growth factor beta 1 ,TGF $\beta$ 1) ; A、F 的交集基因 4 个: 毛状蛋白样 F-肌动蛋白结合蛋白 1 ( coactosin like F-actin binding protein 1 ,COTL1)、聚 ( RC) 结合蛋白 1 ( poly ( rC) binding protein 1 ,PCBP1)、二酰甘油激酶  $\zeta$ ( diacylglycerol kinase zeta ,DGKZ)、接头蛋白 2( docking protein 2 ,DOK2) ,见图 4b。

**2.5 相关基因对胃癌总生存的影响** 上述 12 个交集基因, 导入 KM-plotter, 分析发现: MRPL12、PRR13、COTL1、PCBP1 的高表达与胃癌 OS 延长关联 ( $P < 0.05$ ) ,IDS、LIMD2、PURA、SOLH、CNOT3、TGF $\beta$ 1、DGKZ、DOK2 的高表达与胃癌 OS 降低关联 ( $P < 0.05$ ) ,见表 2 部分基因的生存曲线见图 5。

### 3 讨论

胃癌的综合治疗可提高疗效, 作为综合治疗重

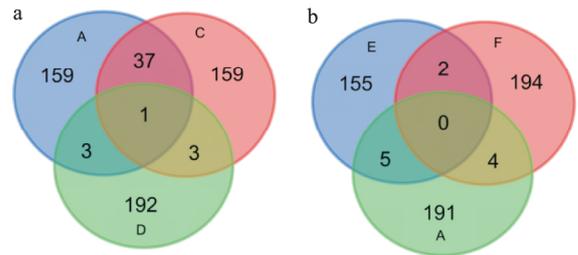


图4 交集基因韦恩图

a: 基因集 A、C、D 的交集; b: 基因集 A、E、F 的交集

表2 交集基因与胃癌总生存的关系

基因名	HR	95% CI	P 值
MRPL 12	0.80	0.67 ~ 0.94	8.4e -03
PRR 13	0.81	0.69 ~ 0.97	1.8e -02
COTL 1	0.70	0.59 ~ 0.83	4.3e -05
PCBP 1	0.56	0.47 ~ 0.67	6.0e -11
IDS	1.36	1.15 ~ 1.62	4.1e -04
LIMD 2	1.55	1.30 ~ 1.84	4.7e -07
PURA	1.72	1.41 ~ 2.10	5.1e -08
SOLH	2.27	1.84 ~ 2.80	2.9e -15
CNOT 3	2.41	1.95 ~ 2.98	1.0e -16
TGF $\beta$ 1	1.53	1.24 ~ 1.90	7.8e -05
DGKZ	2.18	1.80 ~ 2.65	4.4e -16
DOK 2	1.28	1.07 ~ 1.52	6.5e -03

要组成的化疗, 在胃癌手术前后的辅助治疗中发挥重要作用, 明确化疗耐药机制、降低化疗抵抗、提高化疗敏感性, 具有重要的临床意义。顺铂通过多种机制抑制胃癌细胞, 其耐药性的产生可由 PMP22 介导, 但其下游是否与 ISGs 存在关联, 尚无文献报道。为探索 PMP22 下调后减弱胃癌顺铂抵抗是否通过下游的 ISGs 发挥作用, 本研究对 GEO 数据库的 GSE94714 数据集进行了 GSEA 分析, 发现 PMP22 可通过多种 ISGs 相关的下游机制, 对胃癌顺铂抵抗的表型进行调节。

本研究发现, 根据差异倍数 ( fold change , FC) 和统计 P 值进行筛选, PMP22 基因敲减、未敲减组差异表达基因的数目减少, 因此采用未筛选的差异基因, GSEA 分析发现了 NES 前 6 位的 ISGs 条目, 其组成基因用韦恩图方法取交集, 发现了 12 个 ISGs, 均与胃癌的总生存 ( OS) 有关。

随着差异倍数 FC 或 logFC 的增大, 统计分析无差异的基因数逐渐增加, 被排除的基因增多, 筛选出有差异的基因数目逐渐减少, 即上下调基因的数目均减少, 与 Xiao et al<sup>[7]</sup> 进行条件筛选后, 使组间差异表达基因数减少的研究结果一致。预筛选可遗漏部分基因, 造成后续分析无法找到有意义、但是改变细微的重要基因。GSEA 方法可避免遗漏, 较为全

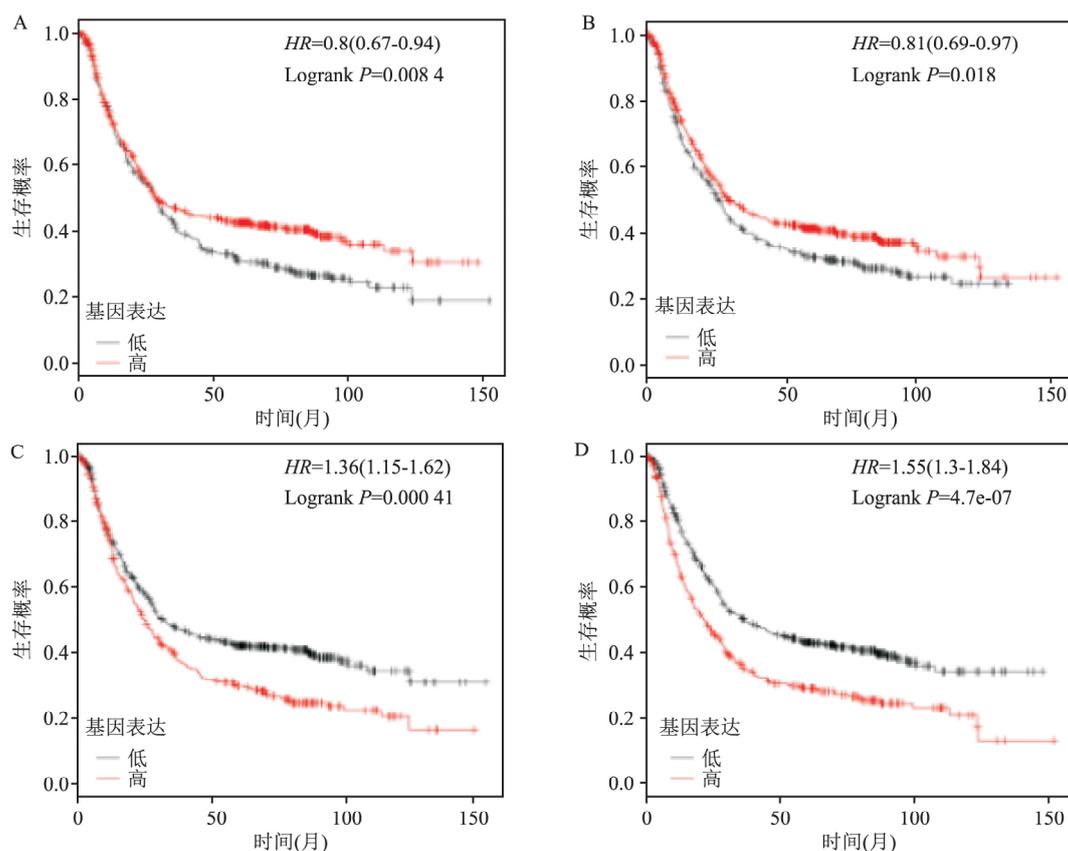


图5 部分交集基因对胃癌 OS 的影响

A: MRPL12 高、低组的 OS 生存曲线; B: PRR13 高、低组的 OS 生存曲线; C: IDS 高、低组的 OS 生存曲线; D: LIMD2 高、低组的 OS 生存曲线

面地分析差异基因,通过预定义基因集,可获得有分类的基因标记信息。笔者通过 GSEA 分析获得的 12 个基因,既与免疫有关,又与胃癌的预后有关,并且在前期进行的常规富集分析中未能被发现<sup>[2]</sup>,表明 GSEA 方法是常规富集分析方法的重要补充。

新发现的 ISGs 通过不同机制参与肿瘤细胞的化疗抵抗和发生发展,部分基因与胃癌化疗抵抗的关系尚未见文献报道。MRPL12 基因可通过改变线粒体代谢、调节细胞的氧耗量,影响肿瘤生长,MR-PL12 敲减后,胰腺肿瘤细胞内的线粒体活性降低,糖酵解加强,肿瘤细胞体外生长减慢,但体内生长加速<sup>[8]</sup>。本研究显示 PRR13 高表达胃癌的 OS 延长,但 Pontikakis et al<sup>[9]</sup> 发现卡铂、紫杉醇治疗后,PRR13 高表达卵巢上皮性癌的 OS 缩短,与化疗抵抗有关,表明 PRR13 表达对不同肿瘤生存时间的作用存在差异,其机制有待进一步阐明。COTL1 基因可增强乳腺癌对化疗的敏感性,与白细胞介素 24、非经典 TGF $\beta$ 1 通路有关<sup>[10]</sup>。PCBP1 可降低豆蔻蛋白表达,增强胃癌腹膜转移细胞对多西他赛的敏感性<sup>[11]</sup>。LIMD2 与甲状腺乳头状癌的转移有关<sup>[12]</sup>。

TGF $\beta$ 1 可能通过细胞黏附、转移相关分子通路影响胃癌预后<sup>[13]</sup>,本研究结果与之相似。DGKZ 作为关键基因,与非小细胞肺癌的进展和预后有关<sup>[14]</sup>。上调因启动子甲基化而沉默的 DOK2 表达,可增强卵巢癌对铂类药物的敏感性<sup>[15]</sup>。IDS、PURA、SOLH、CNOT3 与胃癌化疗抵抗的关系,目前尚未见文献报道,是进一步研究的方向。

本研究的局限性:新发现的 ISGs 与胃癌顺铂抵抗的关系尚需进一步的体内外实验验证,其作用方向和机制依然存在不确定性,初步结果仅为继续研究的基础和线索。

### 参考文献

- [1] Hu S B, Liu C H, Wang X, et al. Pathological evaluation of neoadjuvant chemotherapy in advanced gastric cancer [J]. World J Surg Oncol, 2019, 17(1): 3.
- [2] 汪圣毅,张永红,闫亚飞,等. 胃癌化疗抵抗基因 PMP22 下游的生物信息学机制 [J]. 安徽医科大学学报, 2019, 54(4): 509-14.
- [3] Cai W, Chen G, Luo Q, et al. PMP22 regulates self-renewal and chemoresistance of gastric cancer cells [J]. Mol Cancer Ther, 2017, 16(6): 1187-98.

- [4] Klapdor R, Wang S, Hacker U, et al. Improved killing of ovarian cancer stem cells by combining a novel chimeric antigen receptor-based immunotherapy and chemotherapy [J]. *Hum Gene Ther*, 2017, 28(10): 886–96.
- [5] 王 潇, 尹天舒, 李柏逸, 等. 基因功能富集分析的研究进展 [J]. *中国科学: 生命科学*, 2016, 46(4): 363–73.
- [6] Powers R K, Goodspeed A, Pielke-Lombardo H, et al. GSEA-In-Context: identifying novel and common patterns in expression experiments [J]. *Bioinformatics* 2018, 34(13): i555–64.
- [7] Xiao Y, Feng M, Ran H, et al. Identification of key differentially expressed genes associated with nonsmall cell lung cancer by bioinformatics analyses [J]. *Mol Med Rep*, 2018, 17(5): 6379–86.
- [8] Chen Y, Cairns R, Papandreou I, et al. Oxygen consumption can regulate the growth of tumors, a new perspective on the Warburg effect [J]. *PLoS One*, 2009, 4(9): e7033.
- [9] Pontikakis S, Papadaki C, Tzardi M, et al. Predictive value of ATP7b, BRCA1, BRCA2, PARP1, UIMC1 (RAP80), HOXA9, DAXX, TXN (TRX1), THBS1 (TSP1) and PRR13 (TXR1) genes in patients with epithelial ovarian cancer who received platinum-taxane first-line therapy [J]. *Pharmacogenomics J*, 2017, 17(6): 506–14.
- [10] Xia L, Xiao X, Liu W L, et al. Coactosin-like protein CLP/Cot11 suppresses breast cancer growth through activation of IL-24/PERP and inhibition of non-canonical TGF $\beta$  signaling [J]. *Oncogene*, 2017, 37(3): 323–31.
- [11] Ji F J, Wu Y Y, An Z, et al. Expression of both poly (C) binding protein 1 (PCBP1) and miRNA-3978 is suppressed in peritoneal gastric cancer metastasis [J]. *Sci Rep*, 2017, 7(1): 15488.
- [12] Pinheiro Dos Santos M J C, Bastos A U, da Costa V R, et al. LIMD2 is overexpressed in BRAF V600E-positive papillary thyroid carcinomas and matched lymph node metastases [J]. *Endocr Pathol*, 2018, 29(3): 222–30.
- [13] 刘梦园, 吴 瑛, 孙明军. 公共数据库 GEO 分析 TGFBI1 在胃癌中表达及临床意义 [J]. *临床军医杂志*, 2016, 44(8): 862–5.
- [14] Feng A, Tu Z, Yin B. The effect of HMGB1 on the clinicopathological and prognostic features of non-small cell lung cancer [J]. *Oncotarget*, 2016, 7(15): 20507–19.
- [15] Fang F, Cardenas H, Huang H, et al. Genomic and epigenomic signatures in ovarian cancer associated with resensitization to platinum drugs [J]. *Cancer Res*, 2018, 78(3): 631–44.

## The enrichment analysis of immunologic signature genes associated with cisplatin resistance in gastric cancer

Wang Shengyi, Cheng Yan, Li Xusheng, et al

(Dept of Gastrointestinal Surgery, Dept of General Surgery, The First Affiliated Hospital of Anhui Medical University, Hefei 230022)

**Abstract Objective** To explore the immunologic signature genes (ISGs) associated with cisplatin resistance in gastric cancer based on gene set enrichment analysis (GSEA). **Methods** GSE94714 dataset from GEO database was used, and the differentially expressed genes (DEGs) were analyzed with GEO2R. The effects of conditional gene screening on gene number were observed. GSEA included all DEGs from gastric cancer cells in drug resistant and non-resistant groups. The DEGs were compared with the molecular signatures database (MSigDB), and the obtained ISGs were screened for intersection, and the effects of ISGs on the prognosis of gastric cancer were analyzed by Kaplan Meier Plotter method. **Results** A total of 34 183 DEGs included 12 452 up-regulated and 17 381 down-regulated genes. The increased fold change (FC) added the number of excluded genes. Six entries with the top normalized enrichment score (NES) were identified by GSEA ( $P < 0.01$ ). The intersection ISGs included mitochondrial ribosomal protein L 12 (MRPL12), proline-rich protein 13 (PRR13), coactosin like F-actin binding protein 1 (COTL1), poly (RC) binding protein 1 (PCBP1), iduronate 2-sulfatase (IDS), LIM domain containing 2 (LIMD2), purine rich element binding protein A (PURA), small optic lobes homolog (SOLH), CCR4-NOT transcription complex subunit 3 (CNOT3), transforming growth factor beta 1 (TGF $\beta$ 1), diacylglycerol kinase zeta (DGKZ), and docking protein 2 (DOK2). Twelve ISGs were associated with the overall survival time of gastric cancer, all of which were statistically significant ( $P < 0.05$ ). **Conclusion** The GSEA method can effectively extract the ISGs of cisplatin resistance in gastric cancer. Newly discovered ISGs, as potential targets, can enhance the study of chemoresistant mechanisms in gastric cancer.

**Key words** gastric cancer; chemoresistance; gene set enrichment analysis; GEO; immunologic signature gene; prognosis