

· 原 著 ·

基于生物信息学的未分化甲状腺癌关键发病机制及其潜在干预靶点研究

潘宗富¹, 方琦璐¹, 张轶雯¹, 李 莉², 黄 萍¹

1. 浙江省肿瘤医院药剂科, 浙江 杭州 310022

2. 浙江省淳安县第一人民医院药剂科, 浙江 杭州 311700

[摘要] **目的:**阐明未分化型甲状腺癌(ATC)的分子病理状态,并挖掘其潜在的干预策略。**方法:**利用 GEO 数据库联合 R 语言分析 ATC 组织与正常甲状腺组织差异表达的基因;利用京都基因与基因组百科全书(KEGG)通路数据库和基因本体(GO)数据库对差异表达的基因进行富集和功能注释;基于 STRING 数据库及 Cytoscape 软件构建蛋白质相互作用网络,分析其关键网络节点和基因簇;最后采用 L1000CDS²数据库预测 ATC 的潜在治疗药物。**结果:**共获得 2087 个差异表达基因。与正常甲状腺组织相比,ATC 组织细胞内信号通路及肿瘤微环境均发生显著改变,包括 PI3K-Akt 信号的持续激活、p53 通路的激活、炎症反应、细胞外基质重塑等。蛋白质相互作用网络提示存在 3 个重要的基因簇和 9 个关键节点。将差异表达基因与 L1000CDS²数据库进行比对后发现 22 个能够逆转 ATC 病理状态的潜在化合物。**结论:**本研究揭示了 ATC 发病机制中的关键节点,为 ATC 的治疗提供了潜在的靶点。



[关键词] 甲状腺肿瘤/病理生理学; 计算生物学; 基因表达; 信号传导; 蛋白质类; 基因调控网络

[中图分类号] R736.1 **[文献标志码]** A

Identification of key pathways and drug repurposing for anaplastic thyroid carcinoma by integrated bioinformatics analysis

PAN Zongfu¹, FANG Qilu¹, ZHANG Yiwen¹, LI Li², HUANG Ping¹ (1. Department of Pharmacy, Zhejiang Cancer Hospital, Hangzhou 310022, China; 2. Department of Pharmacy, First People's Hospital of Chun'an, Hangzhou 311700, China)

Corresponding author: HUANG Ping, E-mail: huangping1841@zjcc.org.cn; <https://orcid.org/0000-0002-9598-502X>

收稿日期:2018-01-26 接受日期:2018-02-13

基金项目:浙江省基础公益研究计划(LQ18H160017);浙江省医药卫生科技计划(2017192848,2018245206);杭州市卫生科技计划(2017B56)

第一作者:潘宗富(1989—),男,博士,药师,主要从事内分泌肿瘤发病机制研究;E-mail: panzf@zjcc.org.cn; <https://orcid.org/0000-0002-4054-6660>

通信作者:黄萍(1973—),女,博士,主任药师,主要从事肿瘤药理研究;E-mail: huangping1841@zjcc.org.cn; <https://orcid.org/0000-0002-9598-502X>

[**Abstract**] **Objective:** To identify hub genes and key pathways associated with anaplastic thyroid carcinoma (ATC), and to explore possible intervention strategy. **Methods:** The differentially expressed genes (DEGs) in ATC were identified by Gene Expression Omnibus (GEO) combined with using R language; the pathway enrichment of DEGs were performed by using Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO). The protein-protein interaction (PPI) network of DEGs was constructed by STRING database and visualized by Cytoscape. Furthermore, the hub genes and key nodes were calculated by MCODE. Finally, the drug repurposing was performed by L1000CDS². **Results:** A total of 2087 DEGs were identified. The DEGs were clustered based on functions and pathways with significant enrichment analysis, among which PI3K-Akt signaling pathway, p53 signaling pathway, inflammatory response, extracellular matrix organization were significantly upregulated. The PPI network was constructed and the most significant three modules and nine genes were filtered. Twenty-two potential compounds were repurposed for ATC treatment. **Conclusion:** Using integrated bioinformatics analysis, we have identified hub genes and key pathways in ATC, and provide novel strategy for the treatment of ATC.

[**Key words**] Thyroid neoplasms/physiopathology; Computational biology; Gene expression; Signal transduction; Proteins; Gene regulatory networks

[J Zhejiang Univ (Med Sci), 2018, 47(2):187-193.]

甲状腺癌是常见的内分泌系统恶性肿瘤,根据 2017 年全球癌症统计,甲状腺癌已位列女性恶性肿瘤的第五位,仅次于乳腺癌、肺癌、直肠癌和宫颈癌^[1]。我国 2015 年甲状腺癌发病达 9 万例,发病率较 10 年前上升约 5 倍,严重威胁国民健康^[2]。未分化型甲状腺癌(anaplastic thyroid carcinoma, ATC)是恶性程度最高的甲状腺癌,患者生存期仅为 3~7 个月,远低于分化型甲状腺癌患者的平均生存期。ATC 是一类难治性的恶性甲状腺癌,病情进展快且缺乏有效的治疗靶点。ATC 不仅对放射性碘抵抗,而且采用传统化学药物治疗和放射治疗均不能取得理想效果,患者常死于局部肿瘤扩张导致的窒息或肿瘤远处转移^[3-4]。因此,寻找 ATC 的发生、发展机制及潜在治疗靶点是目前研究的热点与难点。

目前关于 ATC 的机制研究主要聚焦于肿瘤内分子和基因的变化,包括驱动基因的筛选。ATC 具有较高的突变负荷,其中 *BRAF* 及 *RAS* 被认为是 ATC 主要的驱动基因,同时,TP53 突变(>50%)、端粒逆转录酶(TERT)启动子突变(33%~50%)、丝裂原活化蛋白激酶(MAPK)通路改变及人第 10 号染色体缺失的磷酸酶及张力

蛋白同源基因(*PTEN*)突变引起的磷脂酰肌醇 3-激酶(PI3K)-蛋白激酶 B(Akt)-哺乳动物雷帕霉素靶蛋白(mTOR)通路异常激活可导致其基因组不稳定,而 *CTNNB1* 突变(25%~60%)所致 Wnt 通路持续激活则是 ATC 发生上皮间质转化的重要因素^[5-7]。然而,ATC 的分子病理状态仍未完全明确。因此,构建 ATC 发病的分子机制网络,并挖掘其关键节点,将为 ATC 发生、发展机制研究及寻找治疗靶点提供有利证据。

基于上述背景,我们利用生物信息学技术挖掘 ATC 相关基因芯片,分析其差异表达基因并构建蛋白相互作用网络,进而挖掘关键网络节点,以寻找其关键病理机制。最后,基于基因特征导向搜索引擎(L1000CDS²)预测 ATC 潜在的治疗药物或化合物,为 ATC 的治疗提供新的思路。

1 材料与方法

1.1 基因表达谱芯片筛选及数据处理

利用美国国立生物技术信息中心(NCBI)平台下的基因表达综合数据库(Gene Expression Omnibus, GEO)检索含有人源 ATC 样本的芯片,选取含有 ATC 样本最多且通过芯片质量要求的

基因芯片 GSE65144, 芯片质量采用 R 语言 (<http://www.bioconductor.org/>) affyPLM 数据包对数据集进行回归分析判断, 并进行归一化处理。芯片信息: Affymetrix Human Genome U133 Plus 2.0 Array, 平台为 GPL570, 共包含 13 个正常甲状腺组织样本, 12 个 ATC 样本。

1.2 差异表达基因筛选

采用 R 语言 Limma 数据包对 ATC 与正常组织进行差异表达基因筛选, 筛选标准为 $P < 0.01$, $|\log_2| \text{倍数变化}| \geq 1.5$, 并将探针名转化为标准基因名。

1.3 基因功能富集和注释

基于前述所得差异表达基因, 利用京都基因与基因组百科全书 (Kyoto Encyclopedia of Genes and Genomes, KEGG) 通路数据库进行信号通路的富集。同时通过 DAVID 在线软件 (<https://david.ncicrf.gov>), 依据基因本体 (Gene Ontology, GO) 数据库对差异表达基因进行生物学功能注释。

1.4 蛋白质相互作用网络构建及关键节点筛选

基于 STRING 数据库构建蛋白质相互作用网络, 并应用 Cytoscape 3.5.1 进行可视化分析。通过 MCODE (Version 1.4.2, Bader Lab, University of Toronto) 对蛋白质相互作用网络进行关联度分析, 获得蛋白质相互作用网络中关键蛋白质簇和节点蛋白, 并在 Cytoscape 中进行可视化显示。关键节点筛选标准为节点度 (node degree) 大于 10。

1.5 特征基因导向小分子预测

基于 L1000CDS² 数据库, 根据小分子化合物引起的细胞基因谱变化与肿瘤差异表达基因的相似性, 推测小分子逆转肿瘤病理状态的可能性, 并基于 C-SPADE (<http://cspade.fimm.fi/>) 在线工具对小分子化合物进行聚类分析。

2 结果

2.1 差异表达基因筛选结果

共获得 2087 个差异表达基因, 其中变化差异最大的 30 个基因 (15 个上调, 15 个下调) 见图 1。差异表达基因的聚类结

果显示, 12 个 ATC 样本和 13 个正常甲状腺组织样本分别聚为一类, 且 ATC 组织甲状腺功能特异标志物促甲状腺激素受体 (thyrotropin receptor, TSHR)、甲状腺过氧化物酶 (thyroid peroxidase, TPO)、甲状腺球蛋白 (thyroglobulin, TG) 等表达均显著降低, 提示 ATC 失分化状态。

2.2 差异表达基因功能富集和注释结果

通过分析差异表达基因本体的生物进程 (biological process) 发现, ATC 组织上调基因主要富集于细胞外基质重塑、胶原蛋白催化、细胞分裂、细胞黏附、炎症反应、血管新生、染色体黏连、趋化等过程, 而下调的基因主要富集于氧化-还原反应、甲状腺激素合成、外周神经发育、纤毛组装、纤毛形成、声音知觉、鞘脂生物合成、细胞紧密连接组件等生物过程 (图 2A)。进一步对差异表达基因进行 KEGG 通路富集发现, ATC 组织上调基因主要富集于细胞周期、细胞外基质-受体相互作用、PI3K-Akt 信号通路、吞噬体、焦点黏连、蛋白消化及吸收、肿瘤通路、p53 通路、破骨细胞分化、阿米巴病等通路; 下调基因则主要富集于紧密连接, 甲状腺激素合成, β -丙氨酸代谢, 缬氨酸、亮氨酸和异亮氨酸降解, 组氨酸代谢, 醛固酮调控的钠离子重吸收, 精氨酸和脯氨酸代谢, 过氧化物酶体, 胰岛素抵抗, Hippo 信号通路 (图 2B)。

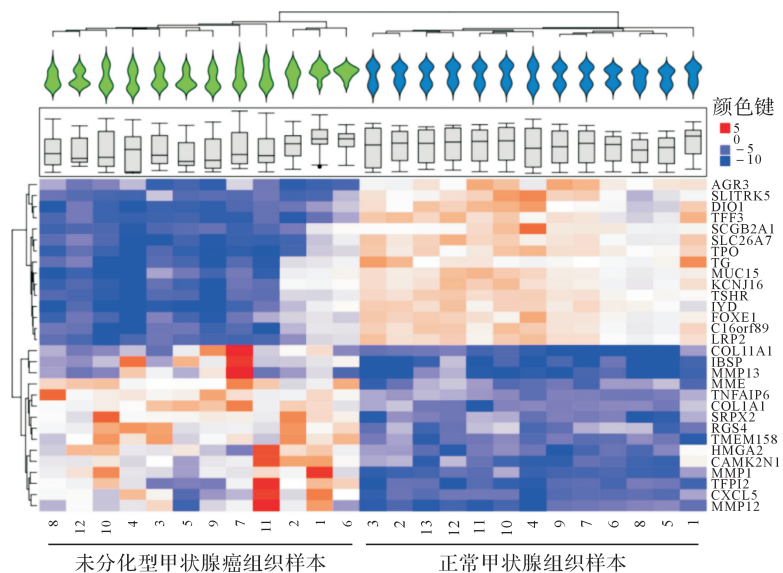


图 1 未分化型甲状腺癌组织与正常甲状腺组织差异表达基因热图和聚类分析结果

Figure 1 Identification and hierarchical clustering analysis of differentially expressed genes between anaplastic thyroid carcinoma and normal thyroid tissues

2.3 蛋白质相互作用网络的构建和关键节点分析
为明确 ATC 的分子病理状态,研究利用差异表达基因构建其潜在的蛋白质相互作用网络,其

中共存在 1295 个网络节点,基因上调以红色表示,下调以青色表示;与其他蛋白质相互作用越多,节点越大;作用越强,连线越粗(图 3)。

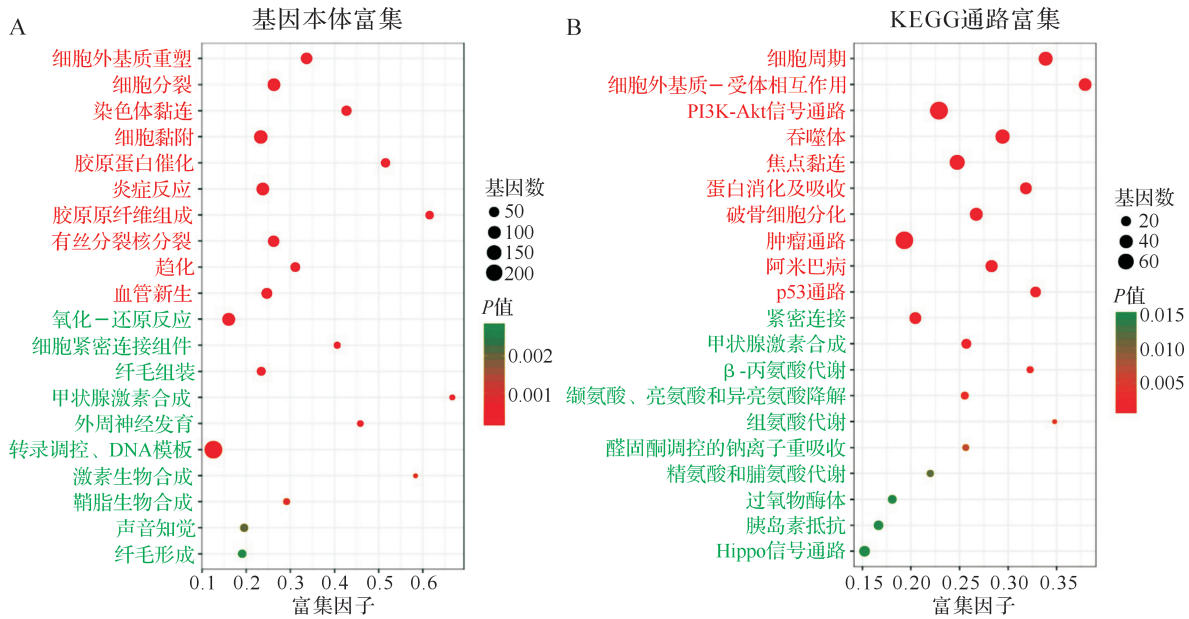
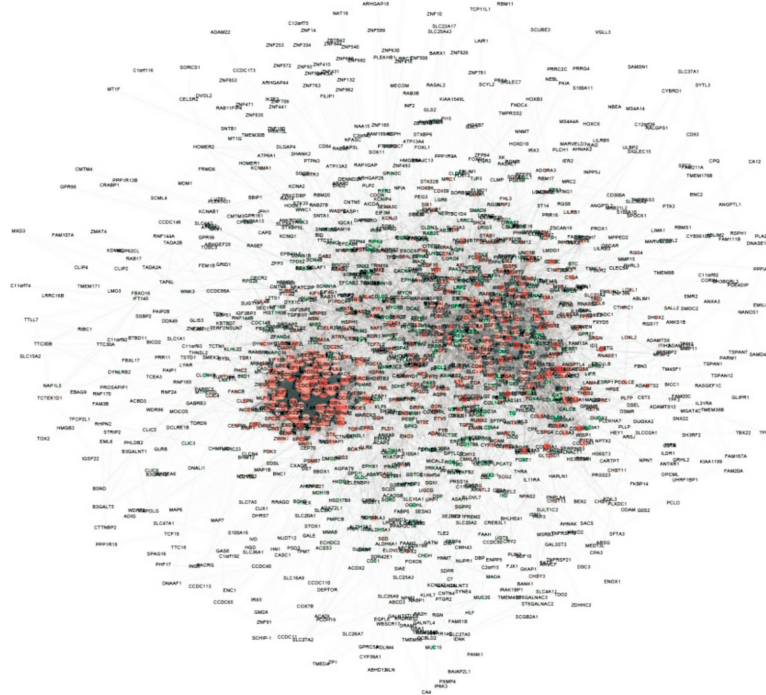


图 2 未分化型甲状腺癌组织与正常甲状腺组织差异表达基因 GO、KEGG 富集和注释结果

Figure 2 GO (Biological Process) terms and KEGG pathways enrichment of differentially expressed genes between anaplastic thyroid carcinoma and normal thyroid tissues



红色为上调基因;青色为下调基因;与其他蛋白质相互作用越多,节点越大,作用越强,连线越粗。

图 3 未分化型甲状腺癌组织与正常甲状腺组织差异表达基因的蛋白质相互作用网络

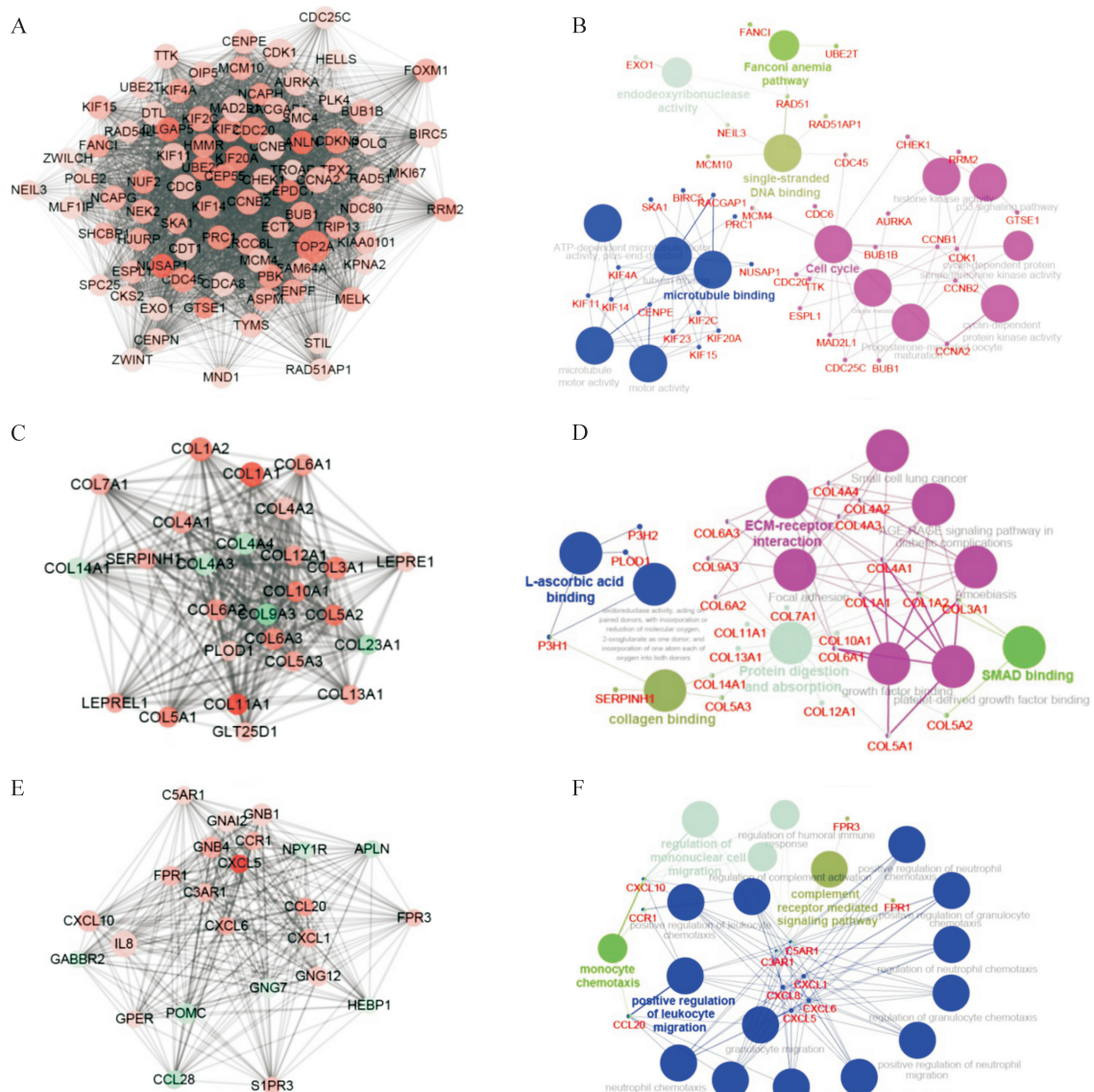
Figure 3 The protein-protein interaction network of differentially expressed genes between anaplastic thyroid carcinoma and normal thyroid tissues

进一步分析网络中的关键节点和基因簇,结果显示,蛋白质相互作用网络中存在三个比较重要的基因簇。第一个基因簇关键节点为 *TOP2A*、*CDK1* 和 *CCNB1*,分别与蛋白质相互作用网络中 204、161、141 个基因存在相互作用(图 4A),该基因簇生物功能主要富集于微管蛋白结合、细胞周期、单链 DNA 结合、范可尼贫血通路及脱氧核糖核酸内切酶活性等通路(图 4B);第二个基因簇关键节点为 *COL1A1*、*COL1A2* 和 *COL3A1*(图 4C),该基因簇功能主要富集于细胞外基质-受体相互作用、SMAD 结合、蛋白消化及吸收、胶原结合及

抗坏血酸结合(图 4D);第三个基因簇关键节点为 *IL-8*、*GNB1* 和 *CXCL10*(图 4E),该基因簇主要与补体受体通路、单核细胞趋化及招募及白细胞迁移调控相关(图 4F)。

2.4 基于差异表达基因的 ATC 干预药物预测

将差异表达基因与 LI000CDS²数据库进行比对后发现 22 个能够逆转 ATC 病理状态的潜在化合物(图 5),其中主要包括酪氨酸激酶抑制剂、表皮生长因子受体抑制剂、MEK 抑制剂、RAF 抑制剂等。这些药物较传统化学药物具有更明确的靶向性,有利于精准治疗。



A、C、E:关键基因簇;B、D、F:基因簇的基因功能注释。

图 4 蛋白质相互作用网络关键基因簇分析及其功能注释

Figure 4 Module analysis of protein-protein interaction network and gene function annotation

- targets for therapy offers new insights for disease treatment[J]. **Ann Surg Oncol**,2007,14(2):719-729.
- [5] LANDA I, IBRAHIMPASIC T, BOUCAI L, et al. Genomic and transcriptomic hallmarks of poorly differentiated and anaplastic thyroid cancers [J]. **J Clin Invest**,2016,126(3):1052-1066.
- [6] SANTARPIA L, EL-NAGGAR A K, COTEG J, et al. Phosphatidylinositol 3-kinase/akt and ras/raf-mitogen-activated protein kinase pathway mutations in anaplastic thyroid cancer [J]. **J Clin Endocrinol Metab**,2008,93(1):278-284.
- [7] LANDA I, GANLY I, CHANT A, et al. Frequent somatic TERT promoter mutations in thyroid cancer; higher prevalence in advanced forms of the disease [J]. **J Clin Endocrinol Metab**, 2013, 98 (9): E1562-E1566.
- [8] HOU P, LIU D, SHAN Y, et al. Genetic alterations and their relationship in the phosphatidylinositol 3-kinase/Akt pathway in thyroid cancer [J]. **Clin Cancer Res**,2007,13(4):1161-1170.
- [9] HUSAIN A, HU N, SADOW P M, et al. Expression of angiogenic switch, cachexia and inflammation factors at the crossroad in undifferentiated thyroid carcinoma with BRAF (V600E) [J]. **Cancer Lett**, 2016,380(2):577-585.
- [10] EJLERSEN B, TUXEN M K, JAKOBSEN E H, et al. Adjuvant cyclophosphamide and docetaxel with or without epirubicin for early TOP2A-normal breast cancer: DBCG 07-READ, an open-label, phase III, randomized trial[J]. **J Clin Oncol**,2017,35(23):2639-2646.
- [11] TURNER J G, DAWSON J L, GRANT S, et al. Treatment of acquired drug resistance in multiple myeloma by combination therapy with XPO1 and topoisomerase II inhibitors [J]. **J Hematol Oncol**, 2016,9(1):73.
- [12] TARPGAARD L S, QVORTRUP C, NYGRD S B, et al. A phase II study of epirubicin in oxaliplatin-resistant patients with metastatic colorectal cancer and TOP2A gene amplification[J]. **BMC Cancer**,2016, 16:91.
- [13] CHRISTNER P J, AYITEY S. Extracellular matrix containing mutated fibrillin-1 (Fbn1) down regulates Col1a1, Col1a2, Col3a1, Col5a1, and Col5a2 mRNA levels in Tsk/+ and Tsk/Tsk embryonic fibroblasts [J]. **Amino Acids**, 2006, 30 (4): 445-451.
- [14] BARCUS C E, O'LEARY K A, BROCKMAN J L, et al. Elevated collagen-I augments tumor progressive signals, intravasation and metastasis of prolactin-induced estrogen receptor alpha positive mammary tumor cells[J]. **Breast Cancer Res**,2017,19(1):9.
- [15] CHEN L, MIN L, WANG X, et al. Loss of RACK1 promotes metastasis of gastric cancer by inducing a miR-302c/IL8 signaling loop [J]. **Cancer Res**, 2015,75(18):3832-3841.
- [16] SANMAMED M F, CARRANZA-RUA O, ALFARO C, et al. Serum interleukin-8 reflects tumor burden and treatment response across malignancies of multiple tissue origins [J]. **Clin Cancer Res**,2014, 20(22):5697-5707.
- [17] KOBAWALA T P, PATEL G H, GAJJAR D R, et al. Clinical utility of serum interleukin-8 and interferon-alpha in thyroid diseases [J]. **J Thyroid Res**,2011,2011:270149.

[本文编辑 余方]

· 读者 · 作者 · 编者 ·

作者投稿时请提供 ORCID

开放研究者与贡献者身份识别码(open researcher and contributor identifier, ORCID)是由汤森路透和自然出版集团等单位于2009年共同发起创建的,其意义与科学文献领域的数字对象标识符(DOI)类似:DOI为科技文献的身份证,一文一证;ORCID为科研人员的学术身份证,一人一证。若尚未获取ORCID的作者请先登录<https://orcid.org/>,注册后免费获取ORCID。本刊从2015年第3期起在作者信息栏添加ORCID,即<https://orcid.org/>后16位数字。如贺晶ORCID为0000-0002-9579-9593,作者信息最后加上<https://orcid.org/0000-0002-9579-9593>。