

基于 TCGA 数据库分析 COL10A1 在 HER2 阳性乳腺癌中的表达及意义

蒋晗旭¹ 王瀚冬² 李义帅³ 于航² 王川⁴ 刘月平¹

(1 河北医科大学第四医院病理科,河北 石家庄 050011; 2 河北医科大学基础医学院;
3 河北省胸科医院胸三科; 4 河北医科大学药理学教研室)

[摘要] **目的** 基于 TCGA 数据库分析 COL10A1 在人表皮生长因子受体 2(HER2)阳性乳腺癌患者中的表达及其与患者预后的关系。**方法** 基于 GEPIA、Oncomine 数据库分析 COL10A1 在乳腺癌中的差异性表达, UALCAN 数据库分析 COL10A1 在不同乳腺癌病理亚型中的表达差异。通过 Kaplan-Meier Plotter 数据库分析 COL10A1 与乳腺癌患者临床预后的关系。**结果** GEPIA、Oncomine 等生物学数据库分析显示,乳腺癌组织中 COL10A1 mRNA 表达水平显著高于正常组织($P<0.01$);HER2 过表达型乳腺癌组织中 COL10A1 表达高于三阴性,差异有统计学意义($P<0.05$)。GEPIA 和 cBioPortal 数据库分析显示,COL10A1 与 HER2 mRNA 表达水平呈正相关($P<0.01$)。Kaplan-Meier Plotter 在线数据库分析显示,HER2 阳性乳腺癌组织中 COL10A1 高表达患者无复发生存率低($P<0.01$)。**结论** COL10A1 在 HER2 阳性乳腺癌组织中呈高表达,并与患者预后呈负相关,有望成为 HER2 阳性乳腺癌患者药物治疗的一个潜在靶点。

[关键词] 乳腺肿瘤;受体,erbB-2;COL10A1 基因;数据库,遗传学;预后
[中图分类号] R737.9;R319.1 **[文献标志码]** A

EXPRESSION AND SIGNIFICANCE OF COL10A1 IN HUMAN EPIDERMAL GROWTH FACTOR RECEPTOR 2-POSITIVE BREAST CANCER; AN ANALYSIS BASED ON TCGA DATABASE JIANG Hanxu, WANG Handong, LI Yishuai, YU Hang, WANG Chuan, LIU Yueping (Department of Pathology, The Fourth Hospital of Hebei Medical University, Shijiazhuang 050011, China)

[ABSTRACT] **Objective** To investigate the expression of COL10A1 in patients with human epidermal growth factor receptor 2 (HER2)-positive breast cancer based on TCGA database and its association with prognosis. **Methods** The differential expression of COL10A1 in breast cancer was analyzed based on the GEPIA and Oncomine databases, and the UALCAN database was used to analyze the difference in the expression of COL10A1 in different subtypes of breast cancer. The Kaplan-Meier Plotter database was used to investigate the association between COL10A1 and clinical prognosis of breast cancer patients. **Results** The analysis based on the GEPIA and Oncomine biological databases showed that the mRNA expression of COL10A1 in breast cancer tissue was significantly higher than that in normal tissue ($P<0.01$), and the expression of COL10A1 in HER2-overexpressing breast cancer was significantly higher than that in triple-negative breast cancer ($P<0.05$). The GEPIA and cBioPortal databases showed a positive correlation between the mRNA expression of COL10A1 and HER2 ($P<0.01$). The Kaplan-Meier Plotter database analysis indicated that the patients with HER2-positive breast cancer and high expression of COL10A1 had a low relapse-free survival rate ($P<0.01$). **Conclusion** Through the in-depth searching of database oncogene information, it is proposed that COL10A1 is highly expressed in HER-positive breast cancer tissue and is negatively correlated with patient's prognosis, and therefore, it may become a potential target for pharmacotherapy for HER-positive breast cancer.

[KEY WORDS] Breast neoplasms; Receptor, ErbB-2; COL10A1 gene; Databases, genetic; Prognosis

乳腺癌是女性中最常见的肿瘤,是女性癌症死亡的主要原因^[1]。据预测,到 2021 年我国 35~49 岁女性中乳腺癌病例将达到 220 万^[2]。在我国乳腺癌患者诊断的平均年龄为 48~49 岁,比西方国家确诊年龄低 10 岁以上^[3]。乳腺癌的治疗效果在很大程度上取决于患者激素受体和人表皮生长因子受体 2(HER2)的表达水平,而乳腺癌的不同的分子亚型

(luminalA 型、luminalB 型、HER2 过表达型及三阴性)预后差异很大^[4]。HER2 是受体酪氨酸激酶(RTK)的 EGFR/ERBB 家族的成员^[5]。HER2 癌基因在大约 25%的乳腺癌中扩增或高表达,并且在大多数癌症中是肿瘤细胞生长的主要驱动力^[6]。临床试验表明,HER2“上癌”是某些肿瘤行为的基础,HER2 作为致癌基因,可促进癌细胞的增殖与转移,并且与肿瘤复发率和患者病死率增高有关^[7]。X 型胶原蛋白(COL10)是结构大分子胶原超家族的一员,由 COL10A1 基因编码^[8]。COL10A1 基因突变与日本型脊椎关节发育不良和 Schmid 型干骺端

[收稿日期] 2019-12-22; [修订日期] 2020-01-23
[基金项目] 国家自然科学基金资助项目(81770407);河北医科大学大学生创新性实验计划项目(USIP2017113,USIP2018048)
[通讯作者] 刘月平,Email:annama@163.com

软骨发育不良有关^[9]。在大多数肿瘤组织中可观察到 COL10A1 高表达,并且其可能与肿瘤血管生成有关。在结直肠癌的早期诊断中,外周血液中 COL10A1 可被视为潜在的生物标志物^[10]。研究显示 COL10A1 在结直肠癌组织中异常表达并参与结直肠癌的进展和 EMT 过程^[11]。此外,通过组织微阵列和基因表达谱分析显示, COL10A1 在乳腺癌中表达上调^[12]。然而, COL10A1 对 HER2 阳性乳腺癌的影响尚不清楚。

肿瘤基因组图谱计划(TCGA)利用基因组测序技术研究 33 种癌症类型及 20 000 多种分子特征。目前 TCGA 已经产生了大量的基因组学、表观基因组学、转录组学和蛋白质组学的大量数据^[13]。本研究深入挖掘基于 TCGA 的肿瘤在线生物信息数据库 cBioPortal、Oncomine、UALCAN 等,综合分析 COL10A1 在 HER2 阳性乳腺癌中的表达及与患者预后的关系,为进一步研究 COL10A1 在乳腺癌治疗中的作用提供一定的参考依据。

1 资料与方法

1.1 利用基因表达谱数据动态分析数据库(GEPIA)筛选乳腺癌高表达基因

通过 GEPIA^[14] (<http://gepia.cancer-pku.cn/>) 搜集相对于正常乳腺组织乳腺癌中高表达的基因,数据集选择“Differential Genes”,Dataset (Cancer name)选择“BRCA”。在方差分析中, $|\log_2 FC|$ 值高于预先设定的阈值, q 值低于预先设定的阈值的基因被认为是差异表达基因。对于前 10 个基因, $|\log_2 FC|$ 值和百分比值高于阈值的基因被认为是过表达基因。

1.2 利用 cBioPortal 数据库分析 COL10A1 基因在乳腺癌中的表达情况

登录 cBioPortal 网站(<http://www.cbioportal.org>)的首页^[15],数据集选择“Breast invasive carcinoma (TCGA, PanCancer Atlas)”,病例组则选择“Tumor Samples with mRNA data (RNA Seq V2) (1100)”,分别分析 COL10A1 的突变、mRNA 水平变化及与 HER2 表达的关系。

1.3 利用 Oncomine 数据库筛选 COL10A1 在不同肿瘤类型中表达的差异

通过 Oncomine 数据库^[16] (<https://www.oncomine.org/>)检索 COL10A1 在不同肿瘤类型中的差异性表达,限定条件为:①基因:COL10A1;②数据类型:mRNA;③显著性: $P < 1E-4$;④差异表达

级别:大于 2 倍;⑤基因排序:前 10%;⑥其他为数据库默认设定。

1.4 利用 UALCAN 数据库分析 COL10A1 在不同类型乳腺癌中的表达

登录 UALCAN 网站(<http://ualcan.path.uab.edu>)首页^[17],数据库类型选择“Breast invasive carcinoma”,输入基因名称“COL10A1”,分析不同类型乳腺癌组织中 COL10A1 基因表达的变化。

数据库类型选择“Metastatic breast cancer [MET500 dataset]”,输入基因名称“COL10A1”,分析 COL10A1 在 HER2 扩增与非扩增样本中表达的差异。

1.5 利用 Kaplan-Meier Plotter 数据分析平台分析 COL10A1 与乳腺癌生存、预后的关系

登录 Kaplan-Meier Plotter 数据分析平台^[18] (<http://kmplot.com/analysis/>),检索 COL10A1 基因,获得 COL10A1 基因表达差异患者的生存曲线。限定条件为无疾病进展生存,“HER2 status”设为“HER2 positive”或“HER2 negative”,对肿瘤病理类型及临床分期、分级等不做限制。

2 结果

2.1 COL10A1 在乳腺癌组织中的表达

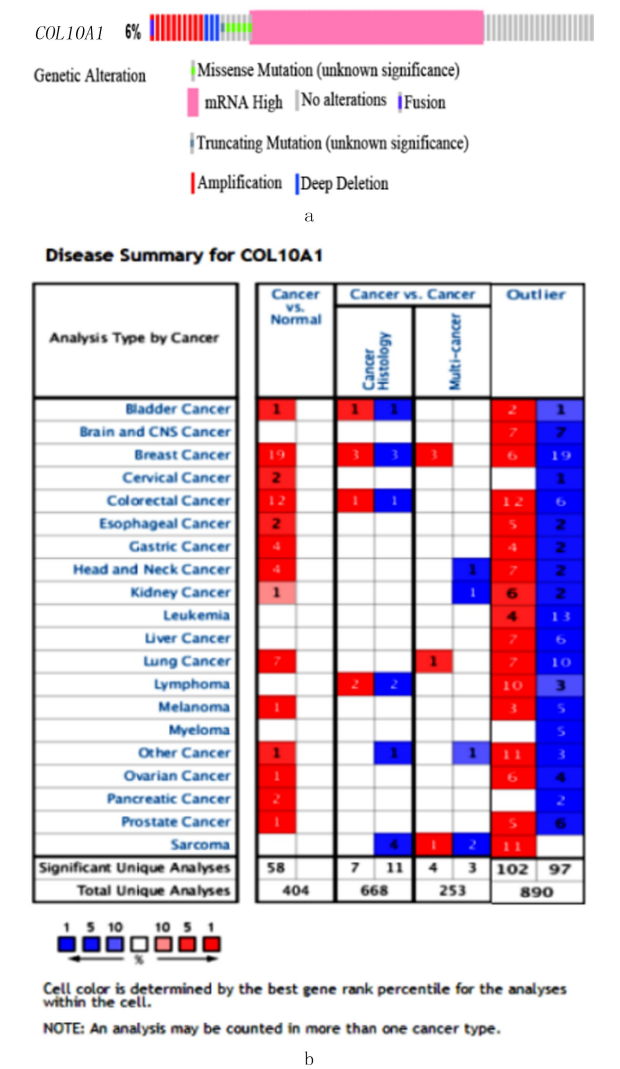
cBioPortal 数据库分析结果显示,994 例乳腺癌患者中,61 例发生了 COL10A1 基因改变,改变的类型包括突变、拷贝数畸变以及 mRNA 水平上调等(图 1a)。GEPIA 以及 Oncomine 数据库分析结果显示, COL10A1 在乳腺癌中呈现高表达(表 1、图 1b)。再利用 UALCAN 平台进一步分析 TCGA 数据库中 COL10A1 在乳腺癌中的差异表达情况,在 TCGA 数据库中总共涉及到正常乳腺组织 114 例,乳腺癌组织 1 097 例, COL10A1 基因在乳腺癌中的表达水平大约为正常乳腺组织的 177.85 倍($P < 1E-12$)。图 2a。

2.2 COL10A1 与 HER2 mRNA 表达的关系

UALCAN 分析显示, luminal 型与 HER2 阳性乳腺癌患者中 COL10A1 表达高于其他类型乳腺癌($P < 0.05$)。见图 2b。且 COL10A1 在 HER2 扩增样本中表达水平明显高于无扩增样本($P < 0.01$)。见图 2c。GEPIA 数据库与 cBioPortal 数据分析显示, COL10A1 与 HER2 表达呈正相关($P < 0.01$)。见图 3a、b。HER2 阳性乳腺癌样本中 COL10A1 表达与 HER2 阴性样本比较,差异具有显著性($P < 0.05$)。见图 3c。

表 1 GEPIA 数据库中乳腺癌组织中高表达的基因

基因	基因 ID	乳腺癌组织(中位数)	正常乳腺组织(中位数)	Log2(倍数变化)	P 值
COL10A1	ENSG00000123500.9	67.46	0.12	5.934	6.52E-08
TFF1	ENSG00000160182.2	133.80	1.27	5.892	2.02E-24
RP11-40C6.2	ENSG00000219928.2	278.05	4.59	5.642	1.98E-85
MMP11	ENSG00000099953.9	167.73	2.56	5.567	9.50E-231
COL11A1	ENSG00000060718.18	37.13	0.50	4.668	1.96E-137
AP000349.2	ENSG00000280178.1	26.15	0.15	4.561	3.61E-121
UBE2C	ENSG00000175063.16	54.58	1.41	4.527	1.12E-227
RP5-940J5.9	ENSG00000269968.1	17.26	0.00	4.191	1.29E-25
AGR3	ENSG00000173467.8	232.90	11.82	4.189	2.97E-24
S100P	ENSG00000163993.6	48.84	1.88	4.113	1.81E-72



a;cBioPortal 数据库分析 COL10A1 在乳腺癌中的突变位点及突变方式;b:Oncomine 数据库中 COL10A1 在不同肿瘤中的差异性表达

图 1 COL10A1 在不同肿瘤中的表达情况及突变方式

2.3 COL10A1 与 HER2 阳性乳腺癌患者预后之间的关系

对 Kaplan-Meier Plotter 数据库分析显示,在全部乳腺癌样本中,COL10A1 与患者无复发生存

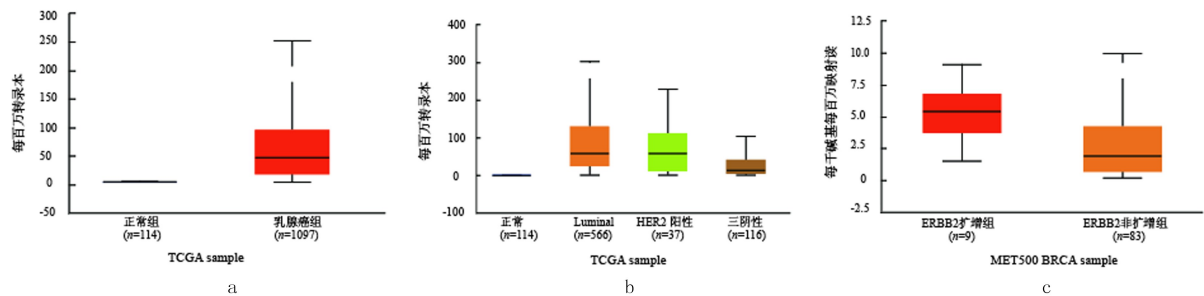
率(RFS)无关。见图 4a。单独筛查 HER2 阴性乳腺癌样本,RFS 也未见明显差异。见图 4b。然而,在 HER2 阳性乳腺癌患者中,相对于 COL10A1 低表达者,COL10A1 高表达者 RFS 降低($P<0.01$)。见图 4c。

3 讨 论

乳腺癌是一种高度异质性的恶性肿瘤,与之相关的分子标志物种类众多^[19]。其中 HER2 表达在乳腺癌进展中发挥着重要作用。尽管目前有针对性 HER2 阳性乳腺癌患者的靶向药物,但抗 HER2 治疗的乳腺癌患者往往会产生耐药性,致使出现新的进展,预后仍不理想^[20],因此本研究期望通过生物信息学数据分析发现 HER2 阳性乳腺癌中的新的靶点。

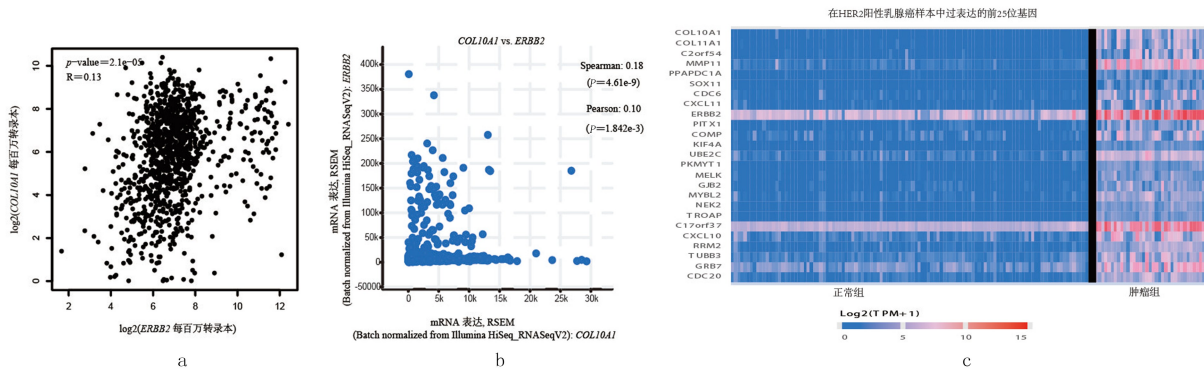
研究显示 COL10A1 在结直肠癌、食管癌和胰腺癌等肿瘤中高表达^[21-23],并且免疫荧光染色表明 COL10A1 定位于乳腺癌的肿瘤脉管系统,而在正常乳腺组织的脉管系统中水平较低^[24]。有研究结合 RNA 表达谱以及免疫组织化学方法分析发现,COL10A1 在对化疗无反应的 ER 阳性/HER2 阳性型乳腺癌样本中表达上调^[25]。但其在 HER2 阳性乳腺癌中的特异性作用尚不清楚。

TCGA 是迄今为止最具实力和最成功的癌症基因组计划之一。TCGA 项目已经生成、分析并提供了超过 1.1 万例代表 30 多种不同癌症类型的基因组和转录组数据。Oncomine 数据库是一个研究肿瘤相关基因的开放数据库,其整合了 TCGA、GEO 和已发表文献来源的 DNA 和 RNA-seq 数据,能帮助人们筛选一些值得研究的靶分子,有利于提高对癌症的诊断、治疗和发生发展机制的深入研究。Ualcan 数据门户是为了促进 TCGA 数据库的基因查询而开发的交互式网络资源,便于分析鉴定肿瘤



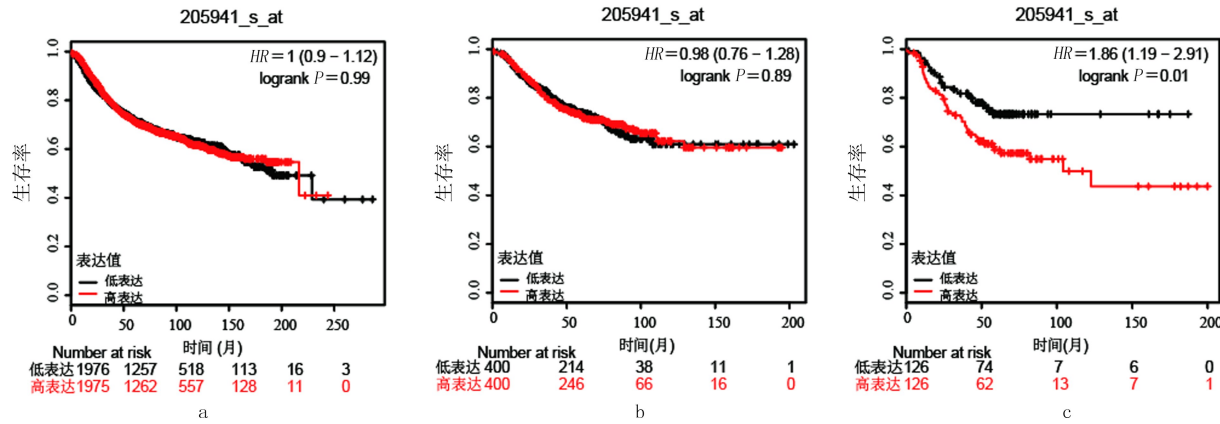
a:COL10A1 在乳腺癌中基于样本类型的表达;b:COL10A1 在乳腺癌亚型中的表达;c:基于 *HER2* 基因扩增亚类的 *COL10A1* 在乳腺癌中的表达

图 2 *COL10A1* mRNA 在 *HER2* 基因扩增样本中的表达情况



a:GEPIA 数据库分析显示,*COL10A1* 与 *HER2* mRNA 表达呈正相关;b:cBioPortal 分析显示,*COL10A1* 与 *HER2* mRNA 表达水平呈正相关;c:UALCAN 分析结果显示,*HER2* 阳性乳腺癌组织中有差异表达的基因

图 3 *COL10A1* 与 *HER2* mRNA 表达水平的关系



a、b、c 分别为 Kaplan-Meier Plotter 数据库中所有乳腺癌样本、*HER2* 阴性乳腺癌样本及 *HER2* 阳性乳腺癌样本

图 4 *COL10A1* 高表达与乳腺癌患者无复发生存率(RFS)的关系

相关的生物分子标记物,从而分析肿瘤发生、进展和转移的相关分子机制,使癌症研究人员更容易获得TCGA 数据和分析结果。此外,cBioPortal 癌症基因组学在线门户网站是一个交互式探索多维癌症基因组学数据集的开放网站,对于肿瘤的诊断、治疗及预后评估具有重要价值。

本研究基于上述肿瘤在线生物信息数据库分析 *COL10A1* 在乳腺癌中的表达及与患者预后的关系。首先通过 GEPIA 及 Oncomine 数据库筛查发现 *COL10A1* 在乳腺癌中高表达。然后利用 UAL-

CAN 进一步分析 TCGA 数据库中 *COL10A1* 在不同类型的乳腺癌中的表达情况,发现 *COL10A1* 在 *HER2* 扩增样本中表达水平明显高于无扩增样本,结果有统计学意义。为了探究 *COL10A1* 与 *HER2* 表达的关系,本研究利用 GEPIA 与 cBioPortal 数据库筛查分析显示,*COL10A1* 与 *HER2* 表达呈显著正相关。以上结果提示 *COL10A1* 在乳腺癌中可能为促癌因素,且与 *HER2* 相互作用,影响乳腺癌的进展。同时 Kaplan-Meier Plotter 数据库分析表明,*COL10A1* 高表达与 *HER2* 阳性乳腺癌患者预

后较差有关,而与 HER2 阴性乳腺癌患者预后无关。以上结果显示 COL10A1 可能与 HER2 相互促进,共同影响乳腺癌患者的预后。

综上所述,通过生物信息数据库综合分析发现, COL10A1 在乳腺癌中高表达,且与 HER2 阳性乳腺癌患者的不良预后相关,可以作为 HER2 阳性的乳腺癌患者临床预后评估的潜在标志物。鉴于 COL10A1 基因表达的种族和个体特异性,后续工作中应在乳腺癌患者肿瘤组织样本中进行验证,并通过基础实验验证靶基因 COL10A1 在 HER2 阳性乳腺癌发生发展中的作用,为 HER2 阳性乳腺癌患者的治疗、诊断和预后提供新的依据。

[参考文献]

[1] SIEGEL R L, MILLER K D, JEMAL A. Cancer statistics[J]. CA Cancer J Clin, 2017,67(1):7-30.

[2] LINOS E, SPANOS D, ROSNER B A, et al. Effects of reproductive and demographic changes on breast cancer incidence in China: A modeling analysis[J]. J Natl Cancer Inst, 2008,100(19):1352-1360.

[3] LI T, MELLO-THOMS C, BRENNAN P C. Descriptive epidemiology of breast cancer in China: Incidence, mortality, survival and prevalence[J]. Breast Cancer Res Treat, 2016, 159(3):395-406.

[4] 李开国,王茹月,杨晓慧,等. 基于 Oncomine 数据库分析 Pygopus2 基因在乳腺癌组织中的表达及其临床意义[J]. 中国癌症防治杂志,10(6):36-40.

[5] 李梅. HER-2 阳性乳腺癌靶向治疗药物研究进展[J]. 实用医药杂志, 2012(8):86-88.

[6] STUHL MILLER T J, MILLER S M, ZAWISTOWSKI J S, et al. Inhibition of lapatinib-induced kinome reprogramming in ERBB2-positive breast cancer by targeting BET family bromodomains[J]. Cell Rep, 2015,11(3):390-404.

[7] SUN B W, MASON S, WILSON R C, et al. Inhibition of the transcriptional kinase CDK7 overcomes therapeutic resistance in HER2-positive breast cancers[J]. Oncogene, 2020,39(1): 50-63.

[8] KIELTY C M, KWAN A P, HOLMES D F, et al. Type X collagen, a product of hypertrophic chondrocytes[J]. Biochem J, 1985,227(2):545-554.

[9] GEBHARD S, HATTORI T, BAUER E, et al. BAC constructs in transgenic reporter mouse lines control efficient and specific LacZ expression in hypertrophic chondrocytes under the complete COL10A1 promoter[J]. Histochem Cell Biol, 2007,127(2):183-194.

[10] SOLÉ X, CROUS-BOU M, CORDERO D, et al. Discovery and validation of new potential biomarkers for early detection of colon cancer[J]. PLoS One, 2014,9(9):e106748.

[11] HUANG H P, LI T T, YE G T, et al. High expression of COL10A1 is associated with poor prognosis in colorectal cancer

[J]. Onco Targets Ther, 2018,11:1571-1581.

[12] MALVIA S, BAGADI S A R, PRADHAN D, et al. Study of gene expression profiles of breast cancers in Indian women[J]. Sci Rep, 2019,9(1):10018.

[13] 吴迪,苏琪,殷红专. 基于 Ualcan 数据库分析 WASH2P 在结肠癌中的表达及预后[J]. 解剖科学进展, 2019,25(6):645-647.

[14] TANG Z F, LI C W, KANG B X, et al. GEPIA: A web server for cancer and normal gene expression profiling and interactive analyses[J]. Nucleic Acids Res, 2017,45(W1):W98-W102.

[15] GAO J J, AKSOY B A, DOGRUSOZ U, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal[J]. Sci Signal, 2013,6(269):p11.

[16] RHODES D R, KALYANA-SUNDARAM S, MAHAVISNO V, et al. Oncomine 3.0: Genes, pathways, and networks in a collection of 18, 000 cancer gene expression profiles[J]. Neoplasia, 2007,9(2):166-180.

[17] CHANDRASHEKAR D S, BASHEL B, BALASUBRAMANYA S A H, et al. UALCAN: A portal for facilitating tumor subgroup gene expression and survival analyses[J]. Neoplasia, 2017,19(8):649-658.

[18] NAGY Á, LÁNCZKY A, MENYHÁRT O, et al. Validation of miRNA prognostic power in hepatocellular carcinoma using expression data of independent datasets[J]. Sci Rep, 2018,8(1):9227

[19] 王西礼. 新辅助化疗对乳腺癌组织中 HER-2、Ki-67 表达的影响及其临床意义[J]. 实用癌症杂志, 2019,34(8):1345-1347.

[20] JIANG N, LIN J J, WANG J, et al. Novel treatment strategies for patients with HER2-positive breast cancer who do not benefit from current targeted therapy drugs[J]. Exp Ther Med, 2018,16(3):2183-2192.

[21] 蒋炳林,何永亮,陈开蓉. MMP3、THBS2、COL10A1 和 miR-1229 对早期结直肠癌的诊断价值[J]. 标记免疫分析与临床, 2019,26(6):994-999.

[22] LI J L, WANG X, ZHENG K, et al. The clinical significance of collagen family gene expression in esophageal squamous cell carcinoma[J]. Peer J, 2019,7:e7705.

[23] 陈况,柳俨哲,王子政,等. X 型胶原蛋白在胰腺癌中的表达及与胰腺癌预后的关系[J]. 解放军医学院学报, 2019,40(4): 373-376,394.

[24] CHAPMAN K B, PRENDES M J, STERNBERG H, et al. COL10A1 expression is elevated in diverse solid tumor types and is associated with tumor vasculature[J]. Future Oncol, 2012,8(8):1031-1040.

[25] BRODSKY A S, XIONG J J, YANG D F, et al. Identification of stromal ColXa1 and tumor-infiltrating lymphocytes as putative predictive markers of neoadjuvant therapy in estrogen receptor-positive/HER2-positive breast cancer[J]. BMC Cancer, 2016,16:274.