

# 基于蛋白质组学的胃腺癌患者预后标志物筛选\*

刘佳璐<sup>1</sup> 韩鑫浩<sup>1</sup> 魏孝礼<sup>2</sup> 陈璐<sup>1</sup> 程丽霞<sup>1</sup> 陈英东<sup>1</sup> 王文晗<sup>1</sup> 张秋菊<sup>1△</sup>

**【摘要】目的** 利用 TCPA 及 UCSC Xena 数据库获取样本信息,构建胃腺癌预后相关蛋白的预测模型,并检验其应用价值,以期为临床治疗与相关机制研究提供参考。**方法** 利用 R 语言软件,通过单因素 Cox 分析和 LASSO-Cox 回归分析筛选与胃腺癌预后有关的蛋白质,并计算风险评分,根据风险评分结果将患者划分为高风险组和低风险组开展生存分析。同时绘制风险评分的 ROC 曲线、蛋白质热图等,用以评估这些蛋白质的预测能力。然后,将风险评分与患者其他临床表型数据结合,进行多因素回归分析,建立预后模型,并绘制列线图。通过 C-index 指标、ROC 曲线、决策曲线等评估预后模型建立的准确性。进行 GO 富集分析与 KEGG 富集分析,探索相关蛋白所参与的生物学过程。**结果** 通过单因素 Cox 分析及 LASSO-Cox 分析筛选出 15 个差异蛋白质 (CKIT、CHK1、CLAUDIN7、COLLAGENVI、DVL3、EGFR\_pY1173、ERALPHA\_pS11、NFKBP65\_pS536、SYK、ETS1、MYOSINIIA\_pS1943、P21、P90RSK、RAPTOR、XBP1),用于计算蛋白质风险评分。通过多因素回归分析,筛选出年龄、性别、M 分期、N 分期、风险评分这 5 种因素,构建预后模型。生存分析结果认为,蛋白质风险评分划分的高低风险组之间的总生存率有显著差异 ( $P < 0.0001$ )。列线图可以用来预测患者 3 年、5 年的生存率,并认为具有较好的预测价值 ( $C\text{-index} = 0.7257$ )。ROC 曲线显示模型在预测中表现出稳定的灵敏度和特异度 ( $AUC = 0.77$  和  $0.86$ )。**结论** 胃腺癌预后相关蛋白质的预测模型可用来预测患者的临床预后,研究结果可为个性化治疗和管理提供参考。

**【关键词】** 胃腺癌 预测模型 预后

**【中图分类号】** R735.2

**【文献标识码】** A

**DOI** 10.11783/j.issn.1002-3674.2025.01.004

## Screening of Prognostic Markers in Patients with Gastric Adenocarcinoma based on Proteomics

Liu Jialu, Han Xinhao, Wei Xiaoli, et al (Department of Health Statistics, School of Public Health, Harbin Medical University, Harbin 150081)

**【Abstract】 Objective** To constructed a predictive model of gastric adenocarcinoma prognosis-related proteins by using TCPA and UCSC Xena databases to obtain sample information. And its application value was tested in order to provide reference for clinical treatment and related mechanism research. **Methods** Using R project, univariate Cox analysis and LASSO-Cox regression analysis were used to select the proteins related to the prognosis of gastric adenocarcinoma, and the risk scores were calculated. According to the risk scores, the patients were divided into high risk group and low risk group for survival analysis. At the same time, draw the ROC curve of the risk score and the protein heat map to evaluate the predictive ability of these proteins. Then, risk scores were combined with other clinical phenotype data of the patients, in order to establish a prognostic model through multivariate regression analysis, and draw a nomogram. The accuracy of the prognostic model was evaluated by C-index, ROC curve and decision curve. GO enrichment analysis and KEGG enrichment analysis were carried out to explore the proteins involved in biological processes. **Results** Fifteen differential proteins (CKIT, CHK1, CLAUDIN7, COLLAGENVI, DVL3, EGFR\_pY1173, ERALPHA\_pS11, NFKBP65\_pS536, SYK, ETS1, MYOSINIIA\_pS1943, P21, P90RSK, RAPTOR, XBP1) were selected by univariate Cox analysis and LASSO-Cox analysis, for calculating the protein risk scores. Through multivariate regression analysis, five factors including age, gender, M stage, N stage and risk score were selected to construct a prognostic model. The results of survival analysis showed that there were significant differences in overall survival between high and low risk groups ( $P < 0.0001$ ). The nomogram can be used to predict 3-year and 5-year survival rate of patients, and it is considered to have good predictive value ( $C\text{-index} = 0.7257$ ). The ROC curve showed that the model exhibited stable sensitivity and specificity in prediction ( $AUC = 0.79$  and  $0.88$ ). **Conclusion** The prognostic model of gastric adenocarcinoma prognosis-related proteins can be used to predict the clinical prognosis of patients, and the research results will help to further guide clinical treatment, which could provide a reference for personalized treatment and management.

**【Key words】** Gastric adenocarcinoma; Prediction model; Prognosis

胃癌 (gastric cancer, GC) 是全球第五大恶性肿瘤,也是全球第三大肿瘤相关的致死因素。它的新发病例数占有癌症的 6%,每年造成约 72 万人死亡,造

成了巨大的经济和社会压力<sup>[1-2]</sup>。2003 年世界卫生组织提出的组织学分型,将胃癌分为腺癌、印戒细胞癌、腺鳞癌、小细胞癌等 9 类<sup>[3]</sup>。其中,胃腺癌 (stomach adenocarcinoma, STAD) 的发生率约占胃恶性肿瘤的 95%。目前,已有 CA724、CA242、CA199 和 CEA 等多种标志物被用于胃癌的筛查与胃癌患者预后的评估,其中 CA724 与胃癌的相关性最高,临床应用广泛<sup>[4-5]</sup>。另外,钙黏蛋白、PCDH7、TPX2、galectin-

\* 基金项目:国家自然科学基金(82073666;82473735);北京医学奖励基金会项目(YXJL-2020-0785-0976)

1.哈尔滨医科大学公共卫生学院卫生统计学教研室(150081)

2.哈尔滨医科大学附属肿瘤医院消化内科

△通信作者:张秋菊,E-mail:qjzh\_81@126.com

8、OLC1、RBMS3、WLS、APJ、PBK/TOPK、CPA4 等组织学指标也被证实对胃癌预后有影响<sup>[6-8]</sup>,但其影响肿瘤复发、转移的分子机制与其临床应用价值尚无定论。研究表明,在乳腺癌、结肠癌、肝癌等肿瘤中,联合多个基因所作的预后模型可显著提高预后的准确性<sup>[9]</sup>。因此,本研究基于胃腺癌预后相关的蛋白表达谱数据和生存数据进行分析,构建蛋白质预后预测模型,探讨其对胃腺癌患者预后的影响,为准确判断预后、指导治疗、改善结局提供更可靠的依据。

## 材料与方法

### 1. 数据获取与样本筛选

本研究数据来源于 2 个公用数据库:从癌症蛋白质组图谱(The Cancer Proteome Atlas, TCPA)数据库(<https://www.tcpaportal.org/>)中下载胃腺癌患者的蛋白质表达数据,包含 392 个临床病例和 218 种蛋白质;从 UCSC Xena 数据库(<https://xena.ucsc.edu/>)中下载相对应的患者生存和临床表型数据,包含 502 个临床病例和 544 个临床表型数据。

从生存数据中提取生存状态、生存时间信息,从临床表型数据中提取年龄、性别、TNM 分期信息,将蛋白表达数据与提取后的患者生存数据、临床表型数据进行整合并剔除有信息缺失的患者。将整合后的数据用于后续患者预后分析。

2. 胃腺癌患者预后蛋白标志物筛选和预后模型的建立

#### (1) 预后相关蛋白质标志物筛选

本研究首先利用单因素 Cox 回归分析筛选与胃腺癌预后有关的蛋白质,考虑到蛋白之间具有多重共线性的问题,再利用 LASSO-Cox 回归进一步过滤标志物。利用模型中蛋白质风险系数与蛋白质表达乘积之和计算蛋白质风险评分(risk score)。

根据预后蛋白质风险评分的加权中位数,将患者划分为高风险组和低风险组<sup>[9]</sup>,对比分析高低风险组患者间生存差异,评价所筛选蛋白质的临床意义,并绘制生存曲线,使结果可视化。同时绘制其 3 年、5 年预后风险评分的受试者工作特征曲线(receiver operation characteristic, ROC)用以评估其预测能力。通常认为,当曲线下面积(area under the curve, AUC)的数值大于 0.60 时,被认为是可以接受的预测;当 AUC 大于 0.75 时,可认为其具有较好的预测能力。本过程通过 R 语言中“survival”包、“survminer”包和“survivalROC”包实现。

绘制蛋白质热图,对比高风险组和低风险组间蛋白质的表达差异。绘制风险曲线,分析风险评分和患

者风险的关联。绘制生存状态图,分析生存状况和患者风险评估的关联。该过程运用了“pheatmap”包、“ggplot2”包来实现。

### (2) 预后模型建立及验证

将风险评分与患者其他临床表型数据结合,拟合 Cox 回归模型,筛选出预后相关危险因素,建立预后模型。

对模型中的预后相关危险因素进行多因素回归分析。绘制列线图(nomogram),用以估计 3 年、5 年、10 年生存率的情形,并计算一致性指数(concordance index, C-index)。通常认为,若 C-index 在 0.7~0.9 左右,代表模型可信度较高。并绘出其 3 年、5 年生存预测的校准曲线,观察预测结果与实际生存率的相符程度。过程通过“rms”包来实现。

绘制其 3 年、5 年生存预测模型的 ROC 曲线用以评估不同时间预测的情况,预测能力通过 AUC 来判断。ROC 曲线的绘制使用了“timeROC”包。

绘制其 1 年、3 年、5 年的决策曲线(Decision Curve Analysis, DCA),在 ROC 曲线的基础上,进一步对模型中的影响因素进行评估。如果模型的曲线远离于参考线,说明该模型具有较高的应用价值。DCA 曲线的绘制使用了“ggDCA”包。

### 3. 蛋白质生物学解释

同样选取 LASSO-Cox 回归中筛选出的蛋白质,分别进行 GO 富集分析与 KEGG 富集分析,探索相关蛋白所参与的生物学过程,并将富集分析结果可视化。富集过程运用“cluster-Profiler”包实现。

## 结果

### 1. 数据资料

经信息整合后,本研究共纳入 288 个临床病例、218 种蛋白质。将可以提取到完整临床数据的 288 例胃腺癌的相关数据进行整理并分析,得到患者的一般情况(表 1)。

表 1 288 例胃腺癌患者的临床特征

临床特征	例数(%)	临床特征	例数(%)
年龄/岁	≤60 岁 104(36.11)	T	T1 8(2.78)
	>60 岁 184(63.89)		T2 61(21.18)
性别	男 192(66.67)	T3	139(48.26)
	女 96(33.33)		T4 80(27.78)
TNM 分期	i 33(11.46)	N	N0 113(39.24)
	ii 95(32.99)		N1 66(22.92)
	iii 132(45.83)		N2 57(19.79)
	iv 28(9.72)		N3 52(18.06)
M	M0 271(94.10)		
	M1 17(5.90)		

2. 胃腺癌患者预后蛋白标志物筛选和预后模型的建立

(1) 预后相关蛋白质标志物筛选

对数据进行单因素 COX 分析,以  $P < 0.05$  为标准,筛选出 CLAUDIN7、CAVEOLIN1、SYK、P90RSK、

MYH11 等 36 个有统计学意义的蛋白质,认为其与胃腺癌预后相关。进一步进行 LASSO-Cox 分析(图 1),筛选出 15 个预后相关的蛋白质,并得到其风险系数(表 2)。

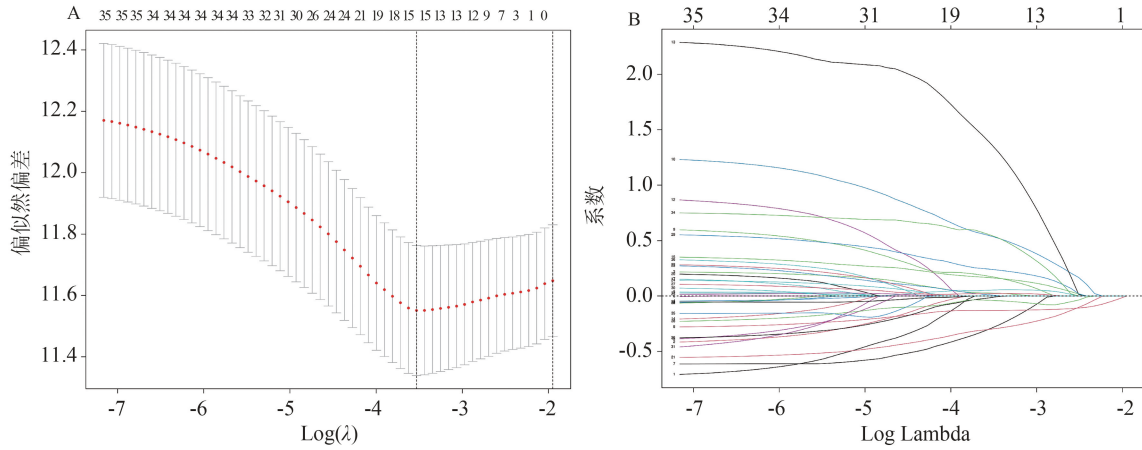


图 1 预后相关蛋白质的 LASSO-Cox 回归分析

表 2 用于胃癌预后相关蛋白模型的蛋白质及风险系数

变量	coef	HR	HR 95%CI	P
CKIT	0.027206637	1.65	1.14~2.38	0.008
CHK1	-0.291762437	0.36	0.14~0.92	0.033
CLAUDIN7	-0.130827285	0.74	0.63~0.86	<0.001
COLLAGENVI	0.175615524	1.81	1.14~2.85	0.011
DVL3	0.540997557	2.37	1.29~4.38	0.006
EGFR_pY1173	0.052885687	1.65	1.13~2.41	0.009
ERALPHA_pS118	1.372035069	4.98	1.09~22.78	0.039
NFKBP65_pS536	0.164683781	1.40	1.09~1.81	0.008
SYK	-0.299175559	0.65	0.49~0.87	0.003
ETS1	0.009333979	1.22	1.01~1.48	0.042
MYOSINIIA_pS1943	-0.014280385	0.73	0.55~0.96	0.025
P21	0.022222447	2.09	1.18~3.69	0.011
P90RSK	-0.039448175	0.52	0.33~0.80	0.003
RAPTOR	0.241143194	2.55	1.26~5.15	0.009
XBP1	0.548397088	2.43	1.12~5.26	0.024

根据各蛋白的风险系数和表达量计算综合风险评分:

风险评分 = (0.027206637 × CKIT 的表达) + (-0.291762437 × CHK1 的表达) + (-0.130827285 × CLAUDIN7 的表达) + (0.175615524 × COLLAGENVI 的表达) + (0.540997557 × DVL3 的表达) + (0.052885687 × EGFR\_pY1173 的表达) + (1.372035069 × ERALPHA\_pS118 的表达) + (0.164683781 × NFKBP65\_pS536 的表达) + (-0.299175559 × SYK 的表达) + (0.009333979 × ETS1 的表达) + (-0.014280385 × MYOSINIIA\_pS1943 的表达) + (0.022222447 × P21 的表达) + (-0.039448175 × P90RSK 的表达) + (0.241143194 × RAPTOR 的表达) +

(0.548397088 × XBP1 的表达)。

按照预后风险评估结果的加权中位数将 288 例患者分成高风险组和低风险组,两组各 144 例。对模型中的 15 个蛋白质分别进行生存分析,结果(图 2)显示,蛋白质 MYOSINIIA\_pS1943、CLAUDIN7、CHK1、SYK 高表达的患者预后良好( $P < 0.05$ ),它们可能是抑癌因子;蛋白质 CKIT、COLLAGENVI、DVL3、EGFR\_pY1173、ERALPHA\_pS118、ETS1、MYOSINIIA\_pS1943、P21、P90RSK、RAPTOR、XBP1 高表达的患者预后不良( $P < 0.05$ ),它们可能是促癌因子。

绘制蛋白质风险评分的生存曲线(图 3A),结果显示  $P < 0.0001$ ,两组间差异显著,低风险组的总生存率明显优于高风险组。

高低风险组间热图(图 3B)显示,蛋白质 SYK、CLAUDIN7、P90RSK、CHK1、MYOSINIIA\_pS1943 在高风险组呈低表达,在低风险组中呈高表达。蛋白质 ERALPHA\_pS118、ETS1、NFKBP65\_pS536、DVL3、RAPTOR、XBP1、EGFR\_pY1173、P21、CKIT、COLLAGENVI 在高风险组呈高表达,在低风险组呈低表达。

风险曲线(图 3C)显示,随着患者风险评分的提高,模型风险评分将相应提高。生存状态图(图 3D)显示,随着患者风险评分的升高,患者的生存状况随之

下降。

胃癌相关蛋白风险评分 3 年、5 年生存的 ROC 曲线(图 4)显示,3 年生存的 AUC 值为 0.705,5 年生存的 AUC 值为 0.727,均大于 0.6,认为其具有一定的预测价值。

### (2) 预后模型建立及验证

将风险评分与患者其他临床表型数据结合,拟合 Cox 回归模型,筛选出预后相关危险因素(表 3),建立预后模型。

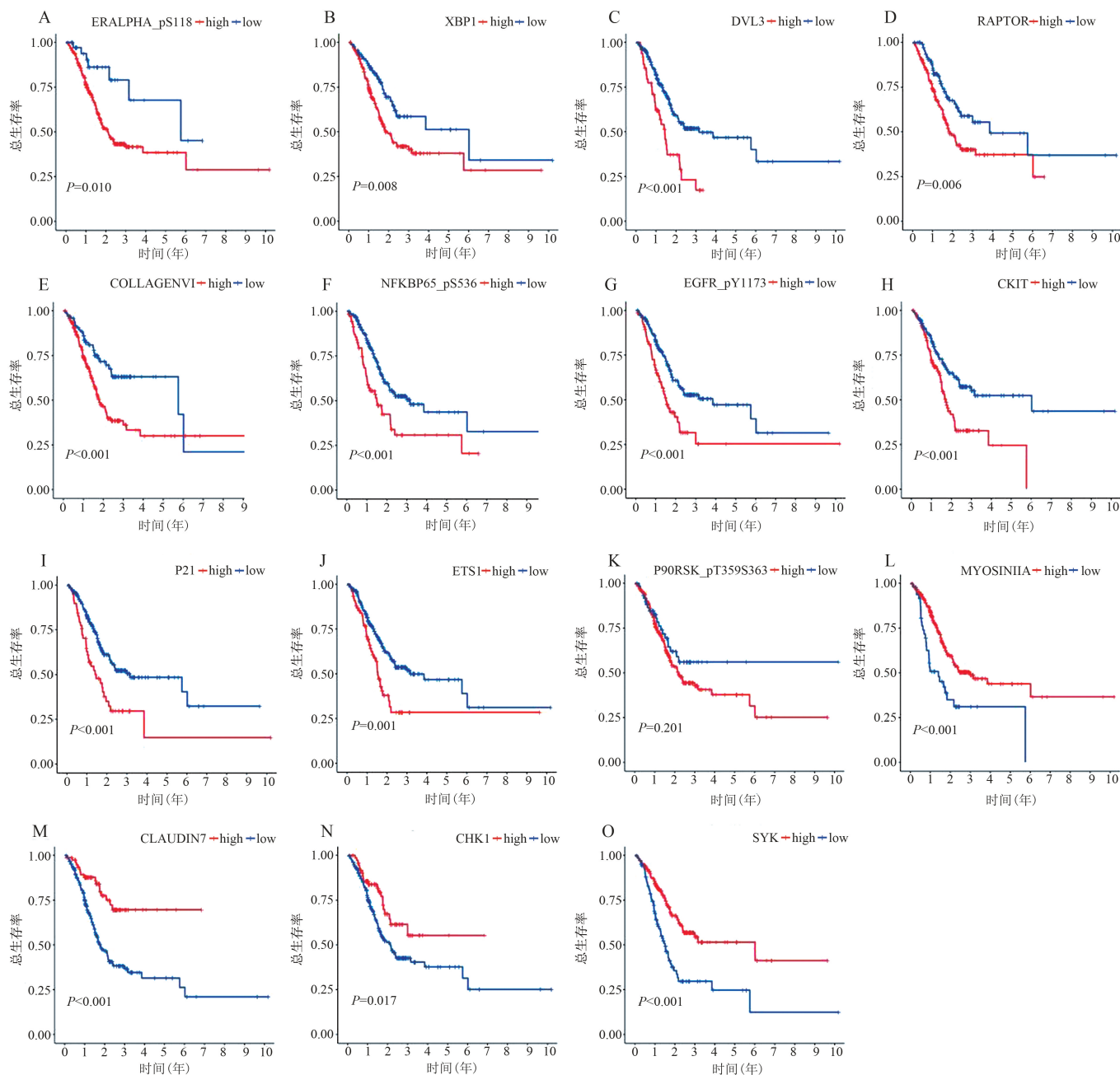


图 2 预测模型各蛋白质的生存分析曲线

列线图,预测 3 年、5 年、10 年存活率的情况(图 5)显示,列线图纳入了年龄、性别、M 分期、N 分期、评分 5 个预后影响因素,将各预后影响因素对应的得分加上,得到的分数向模型下方总分刻度尺划一道垂线,

可以获得患者 3 年、5 年、10 年的存活率。列线图的 C-index 为 0.7257,表明该列线图具有较好的准确性。其 3 年、5 年生存预测的校准曲线(图 6)显示,预测结果与实际生存率较为相符。

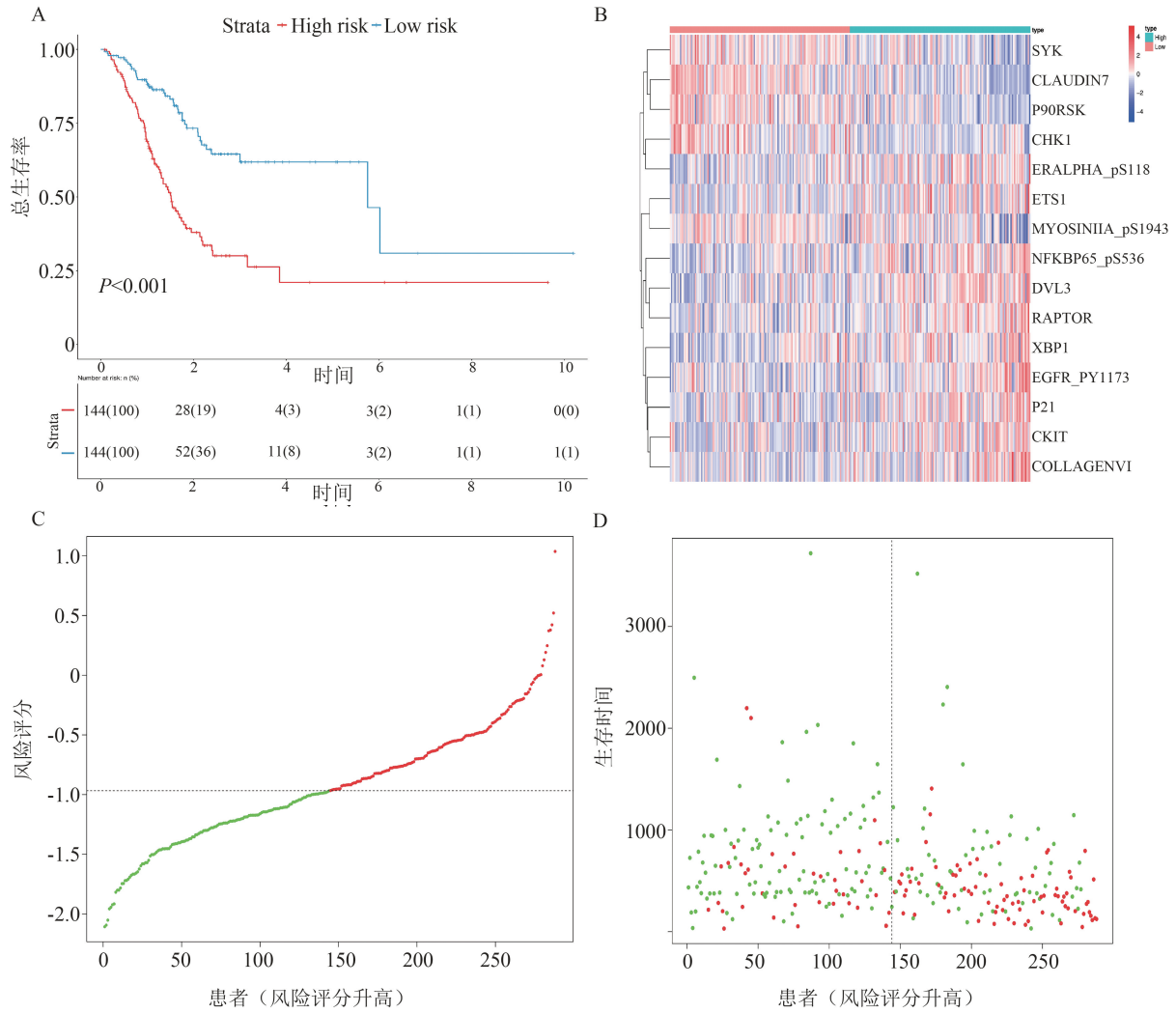


图 3 高风险组与低风险组的生存曲线;风险热图、风险曲线及生存状态图

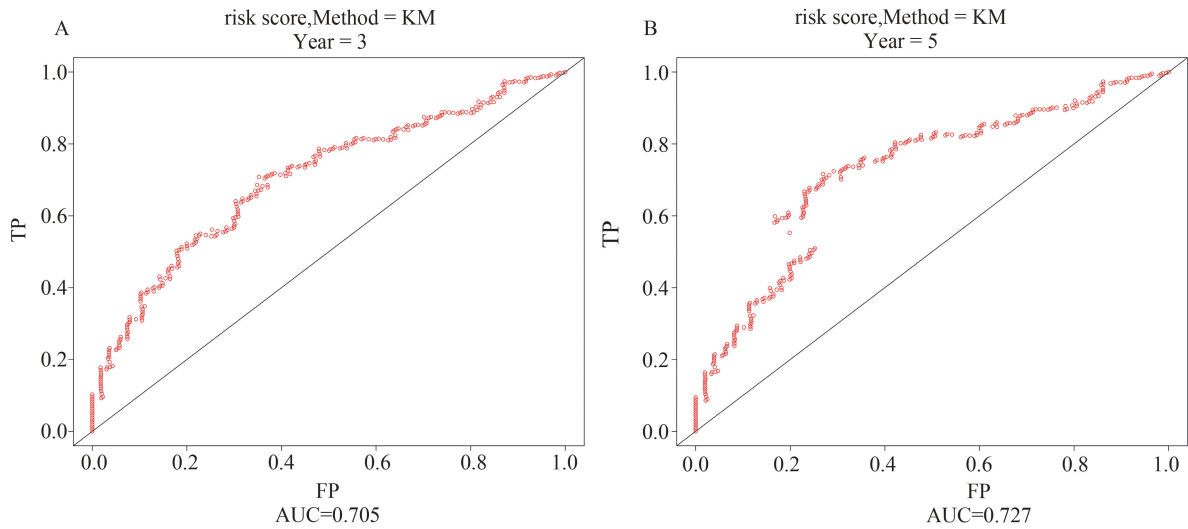


图 4 风险评分的 ROC 曲线

表 3 多因素逐步回归结果

变量	HR	HR. 95L	HR. 95H	P
年龄	1.031	1.011	1.050	0.002
性别	1.528	1.004	2.328	0.048
M 分期	2.265	1.120	4.582	0.023
N 分期	1.358	1.149	1.604	<0.001
评分	4.766	3.144	7.226	<0.001

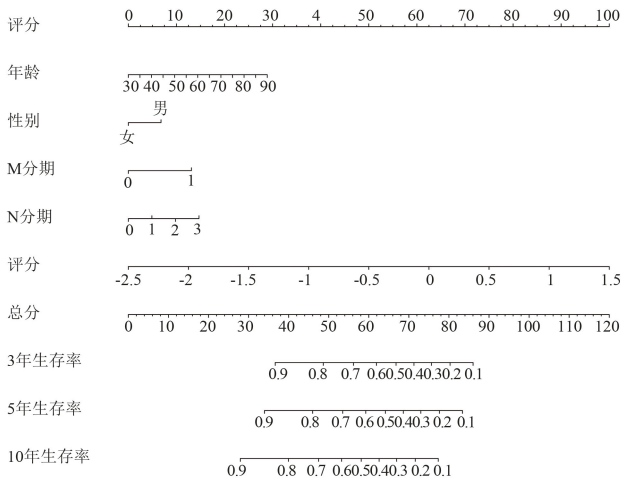


图 5 3年、5年、10年的 nomogram 列线图

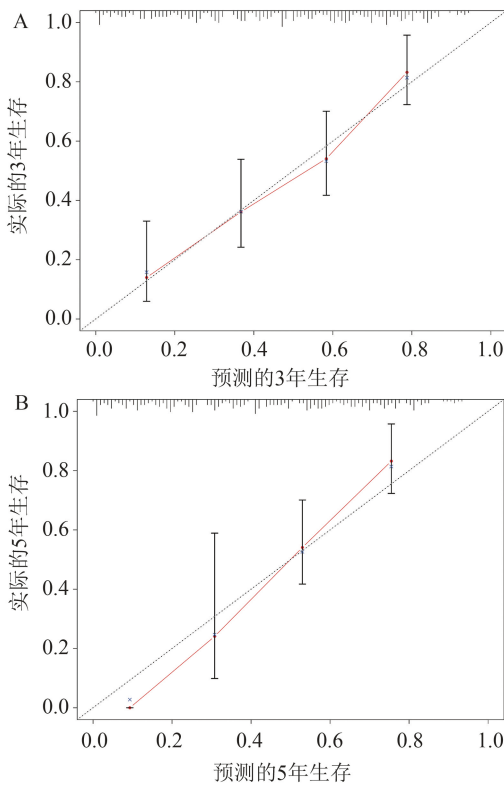


图 6 3年、5年生存率的 nomogram 预测校准图

胃腺癌相关蛋白预后模型3年、5年生存的 ROC 曲线(图 7)显示,所得3年生存的 AUC 值为 0.77,5年生存的 AUC 值为 0.86,均大于 0.75,认为模型具有较好的预测价值。

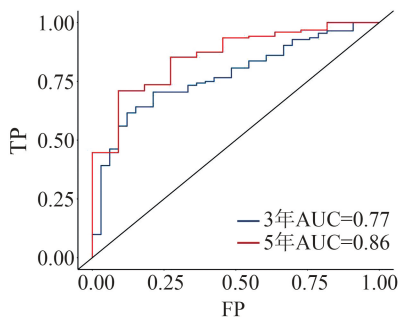


图 7 3年、5年生存的 ROC 曲线

胃腺癌相关蛋白预后模型的 DCA 曲线(图 8)显示,包含蛋白风险评分的总模型的曲线远高于未包含风险得分的参考曲线(浅绿色),说明纳入风险评分后该模型的预测性能得到较大提高。

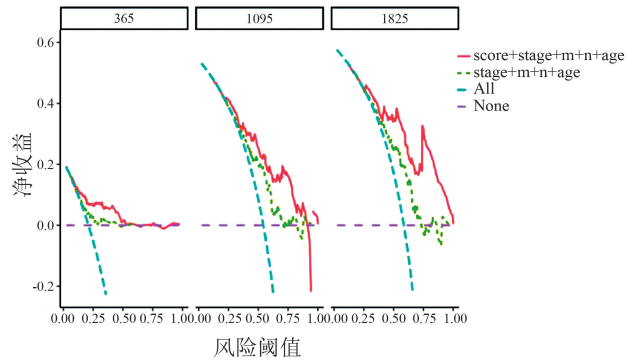


图 8 预后模型 1年、3年、5年的 DCA 曲线

### 3. 蛋白质生物学解释

GO 富集分析结果(图 9A)显示,模型中蛋白质主要参与 B 细胞分化的正向调控、白细胞-细胞黏附的调节、细胞间黏附的调节、造血调节、JUN 激酶活性的调节以及多种蛋白结合的调节等生物学功能(如 Toll 样受体结合、磷脂酶结合、SH2 结构域结合、转录辅助抑制因子结合、磷酸酪氨酸残基结合、蛋白质磷酸化氨基酸结合等)。

KEGG 富集分析结果(图 9B)显示,模型中蛋白质主要参与阿尔兹海默病、神经退行性变、多种信号通路(Notch、Fc epsilon RI、B 细胞受体)、基底细胞癌、肾细胞癌等多种生物学过程。

### 讨论

胃癌是世界上第三大与癌症相关的死亡原因,也是当前最常见的胃肠道恶性肿瘤<sup>[2]</sup>。早期胃癌的临床诊断率低,外科手术是唯一的根治性治疗方法,但有些患者在根治性切除后病情容易复发<sup>[10]</sup>。胃癌的发生发展机制尚不明确,目前对于胃癌预后的判断往往基于临床、病理特性等观察指标,而随着分子生物学的进步,分子标志物对于预后分析的影响也逐渐受到重视,联合这些生物标志物构建预后模型,应用于胃癌的预后分析,能够提高预测的准确性。

雌激素受体  $\alpha$  (estrogen receptor  $\alpha$ , ER $\alpha$ ) 是多种癌症分类和指导治疗的重要生物标志物。pS118-ER $\alpha$  是肽基脯氨酰异构酶 Pin1 的底物,它介导 pS118-P119 键顺反异构化以增强 ER $\alpha$  转录功能<sup>[11]</sup>。本研究中,ERALPHA\_pS118 的风险系数最高,为 1.372,认为其对胃腺癌患者生存影响较为显著,这一点与既往研究一致<sup>[12]</sup>。原因可能是 ER $\alpha$  与生长因子信号转导途径产生串扰作用,能够在不产生配体的情况下刺激 ER $\alpha$ ,进而促使肿瘤生长。肌醇需求酶 1 $\alpha$  (inositol-requiring enzyme-1 $\alpha$ , IRE1 $\alpha$ )-X-box 结合蛋白 1 (X-box binding protein 1, XBP1) 信号传导与癌症的发展

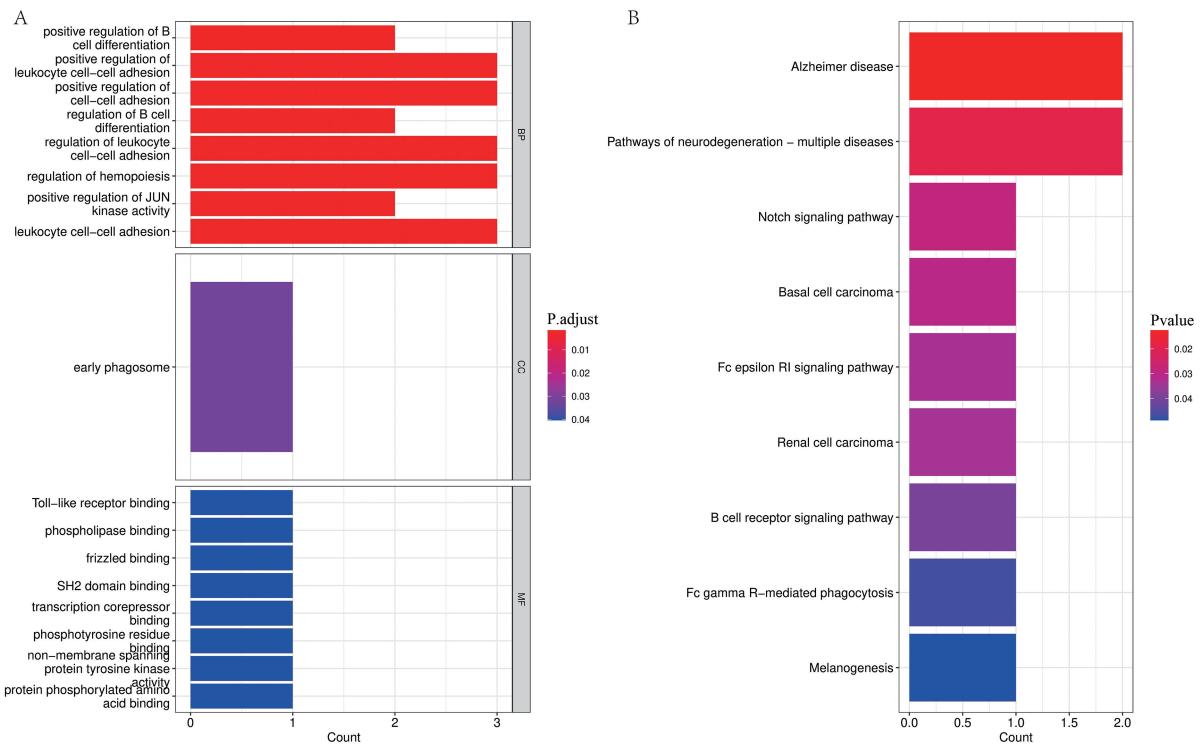


图 9 15 种蛋白质的通路富集分析结果

和进展有关。具有反式激活结构域的活性 XBP1 作为转录调控因子调控细胞下游靶基因的转录表达,与调控细胞生长、凋亡、转移和耐药性中的调节活性促进细胞存活<sup>[13]</sup>。有研究表明,FOXMI、ASPMII 和 DVL3 之间的分子相互作用集中在  $\beta$ -连环素上,能够控制 GC 细胞的 Wnt 活性和干性特性<sup>[14]</sup>。胶原蛋白 VI 在多种癌症的进展中也具有重要意义。胶原蛋白 VI $\alpha$ 3 的过表达已被证明与高级别卵巢癌相关<sup>[15]</sup>。NFK-BP65cyto-RT 是肿瘤局部、远端复发的独立预后因素,该生物标志物高表达的患者更容易出现肿瘤复发<sup>[16]</sup>。幽门螺杆菌诱导的病理生理学可能与 EGFR 信号传导的失调有关。调节 EGFR 信号传导的一个关键机制是配体诱导的内吞作用。幽门螺杆菌在胃上皮细胞长期感染后能够阻断 EGFR 内吞作用和受体降解。抑制通过非受体激酶 c-Abl 的 CagA 依赖性实现,CagA 磷酸化的非依赖性激活反过来又使 EGFR 靶位点 pY1173 磷酸化<sup>[17]</sup>。Ckit 基因位于人染色体 4q12-13,属于原癌基因,受体位于细胞表面,其突变是胃肠道间质瘤发展的早期事件<sup>[18]</sup>。据报道,P21 参与抑制细胞周期蛋白依赖性激酶活性。它可以通过阻断与 G1/S 转换相关的细胞周期蛋白依赖性激酶(cyclin-dependent kinase, CDK)的活性来抑制细胞增殖。HCP5/miR-106b-5p 轴可以转录后调节 p21 表达<sup>[19]</sup>。miR-23b 的下游基因 ETS1 和 TCF4 的表达与 CD4+ T 细胞、CD8+ T 细胞和 B 细胞相关,也是确定的胃癌预后生物标志物<sup>[20]</sup>。

Syk 是一种在造血细胞中广泛表达的非受体型蛋白酪氨酸激酶。Syk 启动子的甲基化与胃癌的发生和

转移有关,是人类胃癌中潜在的抑癌和抗转移基因<sup>[21]</sup>。检查点激酶 1 (cell cycle checkpoint kinase 1, Chk1) 是 DDR 中的关键效应子,是一种通过进化高度保守的多功能 Ser/Thr 激酶蛋白质,是所有细胞周期检查点的关键组成部分。控制 Chk1 调节的主要已知机制是特定残基 Ser317 和 Ser345(均在 C 末端结构域中)的磷酸化,从而导致催化活化。核 Chk1 积累与无进展生存期减少之间存在相关性。Chk1 的表达在转录水平受 p53 和 RB/E2F1 控制,Chk1 的过度表达与某些癌症类型如卵巢癌中的放射治疗抗性相关<sup>[22]</sup>。Claudin7 是胃肠道肿瘤中的 CIC 生物标志物,在 TEX 中回收。棕榈酰化 clid7 有助于 GEM 相关蛋白的募集,包括膜附着的细胞骨架成分和信号分子<sup>[23]</sup>。p90RSK 作为 Ras/Raf/MEK/ERK 信号通路的下游效应子,通过磷酸化不同的底物和调节各种下游信号级联,调节细胞生长、增殖、存活和运动等众多细胞过程。一般来说,RSK1 和 RSK2 被认为通过促进癌细胞的增殖和存活来促进许多肿瘤的进展,而 RSK3 和 4 可能具有一定的抗肿瘤功能<sup>[24]</sup>。肌球蛋白 II 存在于胃壁细胞中,是与细胞活化相关的板状动力学所必需的。当包含肌球蛋白 ATP 酶抑制剂时,肌球蛋白动力学和板状足形成均被取消,这表明肌球蛋白 II 在与壁细胞活化相关的基底外侧膜动力学中起重要作用<sup>[25]</sup>。S1943 上的磷酸化可促进肌球蛋白-II A 细丝在体外的分解,增强表皮细胞生长因子刺激的肿瘤细胞片状足延伸,这对于体外基质降解和体内实验性转移至关重要<sup>[26]</sup>。这些蛋白质均与肿瘤的发生发展有密切关系,但是在胃腺癌中的相关研究有限。

近年来,研究者对胃癌相关的蛋白标志物进行了广泛的研究。人们通过双向凝胶电泳、蛋白质芯片技术、同位素标记相对和绝对定量技术等蛋白质组学技术探索到了一些有效的胃癌潜在标志物,如 GKN1、SLC3A2、PCSK9 等<sup>[27-28]</sup>。Gu 等人<sup>[29]</sup>通过单因素与多因素 Cox 回归分析的方法,证实了肿瘤异常糖蛋白(tumor abnormal protein, TAP)是胃癌早期筛查和预后评估的有效指标。Sun 等人<sup>[30]</sup>通过免疫组化试验与单因素、多因素生存分析方法,证实了 CRM1 与 CDK5 联合表达可为胃癌的总生存提供一个预后模型。本研究在前人单个蛋白研究的基础上,采用了 COX 回归与 LASSO-Cox 回归分析等统计学方法,构建了多蛋白联合表达的预后模型。Zheng 等人<sup>[31]</sup>通过差异表达分析与 Cox 回归分析,构建了一个由三种蛋白(COLLAGEN VI、CD20、TIGAR)组成的预后模型。Xiong 等人<sup>[32]</sup>通过单因素、多因素 Cox 回归分析,构建了一个由 NDRG1\_pT346、SYK、P90RSK、TIGAR 和 XBP1 构成的五蛋白预测模型。本研究构建模型的思路与前两者有相似之处,我们的优势在于,选择了使用 LASSO-Cox 回归来解决多重共线性等问题,模型预测能力的评估结果如 AUC 值等优于前者。且我们绘制了列线图,用以估计患者的生存率。

综上,在本次研究中我们分析了来自 TCPA 数据库与 UCSC Xena 数据库的胃腺癌预后相关蛋白表达数据与生存数据,通过单变量 Cox 回归分析和 LASSO-Cox 分析筛选出 15 个风险差异相关蛋白,并进一步建立了胃腺癌预后相关蛋白的预后风险模型,证实了其能够给患者带来准确的预后。但是,尽管我们做了细致严格的分析,本研究仍存在一些缺陷。由于研究仅采用了 TCPA 数据库与 UCSC Xena 数据库的样本数据进行分析,因此研究结果仅作为初步的理论探索,后续仍需要临床样本数据来进一步验证,以期为胃腺癌的基因分子机制以及预后估计提供新的思路。

### 参 考 文 献

- [ 1 ] Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018; GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA Cancer J Clin, 2018, 68(6): 394-424.
- [ 2 ] Liao F, Guo X, Lu X, et al. A validated survival nomogram for early-onset diffuse gastric cancer[J]. Aging (Albany NY), 2020, 12(13): 13160-13171.
- [ 3 ] Kushima R, Vieth M, Borchard F, et al. Gastric-type well-differentiated adenocarcinoma and pyloric gland adenoma of the stomach [J]. Gastric Cancer, 2006, 9(3): 177-184.
- [ 4 ] 方宏娇,李承慧. 胃癌患者血清 CA199、CA724 及 CEA 水平与病情严重程度及预后的关系[J]. 川北医学院学报, 2022, 37(2): 248-251.
- [ 5 ] Bai H, Deng J, Zhang N, et al. Predictive values of multidetector-row computed tomography combined with serum tumor biomarkers in preoperative lymph node metastasis of gastric cancer [J]. Chin J Cancer Res, 2019, 31(3): 453-462.
- [ 6 ] 张天瑾,陈兆峰,王玉平,等. 肿瘤标志物在胃癌预后判断应用中的研究进展[J]. 胃肠病学和肝病学杂志, 2017, 26(12): 1321-1326.
- [ 7 ] Chen HF, Ma RR, He JY, et al. Protocadherin 7 inhibits cell migration and invasion through E-cadherin in gastric cancer [J]. Tumour Biol, 2017, 39(4): 1010428317697551.
- [ 8 ] Tomii C, Inokuchi M, Takagi Y, et al. TPX2 expression is associated with poor survival in gastric cancer [J]. World J Surg Oncol, 2017, 15(1): 14.
- [ 9 ] 温贺新,史维俊,葛思堂,等. 6 种蛋白质构建的预测模型在结直肠癌预后预测中的作用:基于 TCPA 数据库[J]. 南方医科大学学报, 2021, 41(3): 439-446.
- [ 10 ] Lee HJ, Kim YH, Kim WH, et al. Clinicopathological analysis for recurrence of early gastric cancer [J]. Jpn J Clin Oncol, 2003, 33(5): 209-214.
- [ 11 ] Rajbhandari P, Schalper KA, Solodin NM, et al. Pin1 modulates ER $\alpha$  levels in breast cancer through inhibition of phosphorylation-dependent ubiquitination and degradation [J]. Oncogene, 2014, 33(11): 1438-1447.
- [ 12 ] 韦依依,孙旸,黄素辉,等. IL-6 经 MEK/ERK 途径促进 ER $\alpha$ -Ser118 磷酸化影响卵巢癌细胞他莫昔芬耐药[J]. 免疫学杂志, 2017, 33(4): 294-300.
- [ 13 ] Shi W, Chen Z, Li L, et al. Unravel the molecular mechanism of XBP1 in regulating the biology of cancer cells [J]. J Cancer, 2019, 10(9): 2035-2046.
- [ 14 ] Hsu CC, Liao WY, Chang KY, et al. A multi-mode Wnt- and stemness-regulatory module dictated by FOXM1 and ASPM isoform I in gastric cancer [J]. Gastric Cancer, 2021, 24(3): 624-639.
- [ 15 ] Xie X, Liu X, Zhang Q, et al. Overexpression of collagen VI  $\alpha$ 3 in gastric cancer [J]. Oncol Lett, 2014, 7(5): 1537-1543.
- [ 16 ] Wen Y, Zhao S, Holmqvist A, et al. Predictive Role of Biopsy Based Biomarkers for Radiotherapy Treatment in Rectal Cancer [J]. J Pers Med, 2020, 10(4): 168.
- [ 17 ] Bauer B, Bartfeld S, Meyer TF. H. pylori selectively blocks EGFR endocytosis via the non-receptor kinase c-Abl and CagA [J]. Cell Microbiol, 2009, 11(1): 156-169.
- [ 18 ] 杜春燕. 胃肠道间质瘤临床治疗与基因突变分析研究[D]. 上海:复旦大学, 2007.
- [ 19 ] Chen W, Zhang K, Yang Y, et al. MEF2A-mediated lncRNA HCP5 Inhibits Gastric Cancer Progression via MiR-106b-5p/p21 Axis [J]. Int J Biol Sci, 2021, 17(2): 623-634.
- [ 20 ] Mei D, Qi Y, Xia Y, et al. Microarray profile analysis identifies ETS1 as potential biomarker regulated by miR-23b and modulates TCF4 in gastric cancer [J]. World J Surg Oncol, 2021, 19(1): 311.
- [ 21 ] Wang S, Ding YB, Chen GY, et al. Hypermethylation of Syk gene in promoter region associated with oncogenesis and metastasis of gastric carcinoma [J]. World J Gastroenterol, 2004, 10(12): 1815-1818.
- [ 22 ] Bargiela-Iparraguirre J, Prado-Marchal L, Fernandez-Fuente M, et al. CHK1 expression in Gastric Cancer is modulated by p53 and RB1/E2F1: implications in chemo/radiotherapy response [J]. Sci Rep, 2016(6): 21519.
- [ 23 ] Kyuno D, Bauer N, Schnölzer M, et al. Distinct Origin of Claudin7 in Early Tumor Endosomes Affects Exosome Assembly [J]. Int J Biol Sci, 2019, 15(10): 2224-2239.