

本文引文格式:秦鹏涛,谢先冬,蒋利和.基于单细胞测序初步探索 KRT19 与胃癌患者良性预后的关联性研究[J].右江民族医学院学报,2024,46(6):905-915.

【论著与临床报道】

## 基于单细胞测序初步探索 KRT19 与胃癌患者良性预后的关联性研究

秦鹏涛<sup>1</sup>, 谢先冬<sup>2</sup>, 蒋利和<sup>1</sup>

(1. 右江民族医学院基础医学院, 广西 百色 533000;  
2. 右江民族医学院临床医学院, 广西 百色 533000)

**摘要:**目的 通过单细胞测序技术为主导的生信技术初步探索角蛋白 19(KRT19)作为胃癌发展和预后诊断生物标志物的价值。方法 应用 TCGA 数据库下载胃癌数据集分析基因的差异表达、ROC 分析、免疫浸润分析、GSEA 富集分析、药物敏感分析;收集 GEO 数据库收集的胃癌样本和胃正常组织样本的单细胞测序数据进行去批次、标准化、降维、聚类、注释、拟时序分析、infercnv 恶性预测分析及 CellChat 分析;利用绘图网站 jvenn 绘制韦恩图;使用 KM plotter 网站对 KRT19 等基因进行 KM 生存曲线的绘制;使用 string 网站对 KRT19 基因进行蛋白质互作分析。结果 KRT19 在胃癌组织高表达( $P < 0.001$ ),尤其是在胃癌组织的上皮细胞高表达;KRT19 高表达与胃癌患者预后良好有关( $P < 0.05$ );KRT19 高表达的上皮细胞在胃癌分化末期出现。KRT19 主要通过 MDK(MK)信号通路影响胃癌上皮细胞之间的相互作用;KRT19 高表达的上皮细胞是良性上皮细胞;KRT19 表达与中性粒细胞、浆细胞浸润成正相关,与 CD8<sup>+</sup>T 细胞浸润成负相关;KRT19 与 KRT18、AFP 等基因存在相互作用;KRT19 在角化、核糖体小亚基生物基因等通路发挥作用;高表达的 KRT19 对 5-氟尿嘧啶、乙二醇等药物的药物敏感性较低表达的 KRT19 高。结论 KRT19 可能在分化末期的胃癌上皮细胞高表达,促进胃组织中的良性上皮细胞出现或增殖,对胃癌的发展具有良好预后的作用。

**关键词:** 胃肿瘤;单细胞测序分析;预后;角蛋白 19

中图分类号:R735.2 文献标识码:A 文章编号:1001-5817(2024)06-0905-11

doi:10.3969/j.issn.1001-5817.2024.06.013

### Preliminary exploration of the association between KRT19 and favorable prognosis in gastric cancer patients based on single-cell sequencing

QIN Pengtao<sup>1</sup>, XIE Xiandong<sup>2</sup>, JIANG Lihe<sup>1</sup>

(1. School of Basic Medicine, Youjiang Medical University for Nationalities, Baise 533000, Guangxi, China; 2. School of Clinical Medicine, Youjiang Medical University for Nationalities, Baise 533000, Guangxi, China)

**Abstract:** **Objective** To initially explore the value of keratin 19 (KRT19) as a biomarker for gastric cancer progression and prognostic diagnosis using single-cell sequencing technology-driven bioinformatics techniques. **Methods** The TCGA database was used to download gastric cancer datasets for differential gene expression analysis, ROC analysis, immune infiltration analysis, GSEA enrichment analysis, and drug sensitivity analysis. Single-cell sequencing data of gastric cancer samples and normal gastric tissue samples collected from the GEO database were subjected to batch correction, normalization, dimensionality reduction, clustering, annotation, pseudotime analysis, infercnv malignant prediction analysis, and CellChat analysis. Venn diagrams

基金项目:右江民族医学院高层次引进人才项目(Y Y2021SK02)

第一作者:秦鹏涛,大学生, E-mail:2581678308@qq.com

通讯作者:蒋利和,博士,教授,研究方向:肿瘤生物学, E-mail:Jianglihe@ymun.edu.cn

were drawn using the jvenn plotting website. Genes including KRT19 were subjected to KM survival curve plotting using the KM plotter website. Protein-protein interaction analysis of the KRT19 gene was performed using the string website. **Results** KRT19 was highly expressed in gastric cancer tissues ( $P < 0.001$ ), particularly in the epithelial cells of these tissues. High KRT19 expression was associated with favorable prognosis in gastric cancer patients ( $P < 0.05$ ). High-expressing KRT19 epithelial cells emerged at the late stage of gastric cancer differentiation. KRT19 primarily influenced interactions between gastric cancer epithelial cells through the MDK (MK) signaling pathway. High-expressing KRT19 epithelial cells were benign. KRT19 expression positively correlated with neutrophil and plasma cell infiltration and negatively correlated with CD8<sup>+</sup> T-cell infiltration. KRT19 interacted with genes such as KRT18 and AFP. KRT19 functioned in pathways such as keratinization and ribosomal small subunit biogenesis. High KRT19 expression exhibited lower drug sensitivity to drugs like 5-fluorouracil and ethylene glycol compared to low KRT19 expression. **Conclusion** KRT19 may be highly expressed in late-stage differentiated gastric cancer epithelial cells, promoting the appearance or proliferation of benign epithelial cells in gastric tissue and contributing to a favorable prognosis in gastric cancer development.

**Key words:** gastric tumors; single-cell sequencing analysis; prognosis; keratin 19

胃癌是全球人类消化系统最常见的恶性肿瘤。通过对胃癌的广泛研究,其发病率和死亡率在近年来均有所下降。然而,根据国际癌症研究机构的数据,胃癌的 5 年发病率为 5.6%,排名第五,死亡率为 7.7%,排名第四<sup>[1]</sup>。目前,胃癌的治疗不仅限于传统的手术切除,还包括放射治疗和化疗等其他治疗方法,以控制局部疾病并提高生存率<sup>[2]</sup>。早期胃癌通常表现为非特异性症状,因此难以被及时发现。因此,大多数患者在首次诊断时被诊断为晚期胃癌<sup>[3]</sup>。探索胃癌的新的预后标志物可以促进对胃癌患者及时实施适当的治疗,从而提高其生存率。

角蛋白 19(KRT19)属于酸性 I 型角蛋白家族的成员,被称为家族中最小的成员(40 kDa)。KRT19 具有高度保守的  $\alpha$ -螺旋中央结构域,对于中间丝的形成起着关键作用。此外,它缺乏非螺旋 C-末端尾域<sup>[4]</sup>。作为细胞质中间丝蛋白,KRT19 可能负责结构刚性和多功能支架,使其成为上皮组织或细胞的潜在生物标志物<sup>[5]</sup>。目前的研究表明,KRT19 可能与多种信号转导因子相互作用,包括受体、激酶、效应器和适配器<sup>[6]</sup>。虽然 KRT19 的预后意义在各种癌症中已经广泛讨论,但其在胃癌中的具体作用尚未被深入探索。

为了探究 KRT19 基因表达与胃癌发展进程的相关性,本研究基于单细胞测序围绕 KRT19 高表达的胃癌组织上皮细胞群进行拟时序分析、恶性预测分析、细胞通讯通路分析,并通过 TCGA 数据库所提供的数据进行胃癌中 KRT19 的基因表达水平分析以及对 KRT19 基因的免疫细胞浸润特征、GSEA 富集分析和药物敏感性进行研究。通过生存曲线分析确定了 KRT19 的预后方向,并通过蛋白质互作分析,探讨了与 KRT19 相互作用的基因。此外,分析了这些基因

相互作用过程中可能对胃癌发展进程的潜在影响。本研究通过上述分析,初步探索了 KRT19 基因影响胃癌发展进程的机制,有助于进一步判断胃癌患者的预后状况。

## 1 材料与方法

1.1 数据准备 从 The Cancer Genome Atlas Program (TCGA, <https://www.cancer.gov/ccg/research/genome-sequencing/tcga>) 中收集胃癌转录组数据和胃癌临床数据,其中有 377 个胃癌组织样本和 34 个正常胃组织样本。并将转录组测序数据转化为 TPM(transcripts permillion mapped reads)格式,用于后续分析。从 Gene Expression Omnibus (GEO, <https://www.ncbi.nlm.nih.gov/geo/>) 中收集了胃癌单细胞测序数据分别来源于数据集 GSE167297 与 GSE184198,其中有 11 个胃癌组织样本和 5 个正常胃组织样本。使用 Seurat 软件包将 GSE167297 和 GSE184198 收集的胃癌单细胞数据进行预处理,区分为正常组和胃癌组。随后,将两组数据分别合并,用于后续分析。

1.2 TCGA 差异分析 通过 R 包 stats 使用基本的统计检验函数 wilcox.test 函数对 TCGA 转录组数据中的胃癌和癌旁数据进行基因差异表达分析,wilcox.test 函数是对两个样本的观测值进行秩和排序,然后计算秩和之和。通过对秩和之和进行假设检验,可以判断两个样本的中位数是否存在显著差异。随后使用 R 包 ggplot2 将结果进行可视化,并筛选出具有差异且在胃癌中呈现高表达的基因(筛选条件:  $FDR < 0.05, \log_2FC > 1$ )。

1.3 单细胞分析 使用 R 包 Seurat 通过 NormalizeData 函数对单细胞数据进行标准化,去除了低质量

的细胞和基因。利用 FindVariableFeatures 函数识别高变基因,并筛选前 2000 个高变基因进行 PCA 分析,利用 RunPCA 函数和 RunTSNE 函数对 2000 个高变基因进行降维,利用 Harmony 函数去除批次效应。筛选  $P < 0.05$  的 PC 进行分群聚类,根据 PCA 的方差解释率和细胞聚类效果,选择了合适的主成分个数,作为后续分析的输入。使用 FindCluster 函数对细胞进行了分群聚类,根据模块化系数和轮廓系数,选择了合适的分辨率参数,得到了最合适的细胞聚类结果,并展示每个聚类前 5 个基因的热图。使用 Seurat 包根据常见的标记基因 epithelial cell (epithelial) (EP-CAM<sup>[7]</sup>, KRT19, CLDN4<sup>[8-9]</sup>), stromal cell (stromal) (PECAM1<sup>[10]</sup>, CLO1A2, VWF), proliferative cell (proliferative) (MKI67, STMN1, PCNA<sup>[11]</sup>), T cell (T) (CD3D, CD3E, CD2), B cell (B) (CD79A<sup>[12]</sup>, IGHG1, MS4A1), myeloid cell (myeloid) (CSF1R<sup>[13]</sup>, CSF3R, CD68) 对细胞群进行标记,对无法标记的细胞群进行剔除。

1.4 筛选目标基因 分别选取单细胞正常组和胃癌组,在注释后标记为 Epithelial cells 的细胞群中 avg\_log2FC 值由大到小的前一百个基因,与差异分析中呈现高表达的基因,导入到韦恩图在线绘制网站(jvenn, <https://jvenn.toulouse.inrae.fr/app/example.html>)。最后根据韦恩图结果,选取在单细胞正常组、胃癌组与差异分析高表达结果里面都含有的基因,目的是为了选出在胃癌组具有更高表达的特征基因。随后在 Kaplan-Meier Plotter (KM plotter, <https://km-plot.com/analysis>) 进行 KM 曲线绘制筛选出胃癌患者中可能具有保护因素的基因(筛选条件:HR<1,  $P < 0.05$ )。选择 KRT19 作为本研究的目标基因。

1.5 受试者工作特征曲线(ROC)判断 KRT19 对胃癌的诊断价值 使用了 pROC<sup>[1.18.4]</sup>包,对 411 个 TCGA 样本(其中有 377 个胃癌组织样本和 34 个正常胃组织样本)进行了 ROC 分析,并用 ggplot2 包绘制了 ROC 曲线。

1.6 对 KRT19 进行表达量分析与单细胞分析 通过 R 语言 boxplot 函数将目标基因 KRT19 在胃癌差异表达情况进行了可视化。通过 R 语言 FeaturePlot 函数分别绘制出目标基因 KRT19 在正常组和胃癌组的细胞群表达情况。使用 R 包 monocle 其中的 plot\_cell\_trajectory 函数绘出 KRT19 在胃癌组拟时序中的表达情况,以此来揭示 KRT19 在细胞状态和发育过程的影响。拟时序结果显示 KRT19 在上皮细胞群具有较高的表达,且 KRT19 在接近分化末期时表达程度变高。因此将胃癌组上皮细胞群单独分出聚类,根据 KRT19 在上皮细胞群的表达量,分别注释为

KRT19 高表达组(KRT19\_high)和 KRT19 低表达组(KRT19\_low)。随后使用 R 包 infercnv 中 infercnv:run 函数对 KRT19 高表达组和 KRT19 低表达组进行恶性预测。最后使用 R 包 CellChat 分析细胞间的通讯网络。

1.7 免疫浸润分析 利用 CIBERSORT 包对 TCGA 的胃癌数据进行免疫浸润分析,利用箱线图查看免疫细胞占比,利用散点图查看目标基因与免疫细胞之间的相关性。

1.8 蛋白质互作分析 SRTING 数据库([https://www.string-db.org/cgi/input?sessionId=bAoFN-SINkoC6&input\\_page\\_show\\_search=off](https://www.string-db.org/cgi/input?sessionId=bAoFN-SINkoC6&input_page_show_search=off))是分析蛋白质相互作用的数据库。通过该数据库构建了 KRT19 的蛋白质-蛋白质相互作用(PPI)网络,设置的最低互交分数为 0.85。

1.9 GSEA 富集分析 首先通过使用 org.Hs.eg.db 包进行了 ID 转化,然后通过 clusterProfiler 包对目标基因进行了 GO 功能和 KEGG 通路的 GSEA 富集分析,并利用 enrichplot 绘制排名前 5 的富集结果(GO 功能和 KEGG 通路的基因集均来自 GSEA 数据库 <https://www.gsea-msigdb.org/gsea/index.jsp>)。

1.10 药物敏感性分析 使用 R 包 oncoPredict 的 calcPhenotype 函数进行药物敏感性评分,其中使用 OSF 平台下载有关 oncoPredict 包的最新整理数据(<https://osf.io/temyk>),下载的数据用作训练数据,以此来构建可以通过学习表达数据与表型之间的关联,从而预测新样本表型的模型。导入 TPM 格式的转录组测序数据作为测试数据并设置参数 batchCorrect='eb' 来批次效应校正后进行药物敏感性评分。按照 KRT19 表达水平的中位值进行高低表达水平的分组,使用 wilcox.test 函数进行差异分析。随后筛选对 KRT19 高表达组更加敏感的药物(筛选条件:  $P < 0.05, \log_2FC < 0$ ),利用 boxplot 函数对筛选后的结果进行可视化。

1.11 统计学方法 本研究采用 R 软件 4.3.0 版本进行数据处理、统计分析和数据可视化。通过 Student's *t* 检验或 Wilcoxon 检验评估两组之间的比较。 $P < 0.05$  被认为差异有统计学意义。

## 2 结果

2.1 TCGA 差异分析结果 根据 TCGA 临床信息将 TCGA 数据分为胃癌组和正常组,并利用 wilcox.test 函数进行基因差异表达分析,使用  $|\log_2FC| > 1$  和  $P < 0.05$  作为差异表达基因的筛选阈值,从而得到 1880 个 DEGs(见图 1)。其中上调表达的基因有 967 个(筛选上调表达的条件:  $FDR < 0.05, \log_2FC > 1$ )。

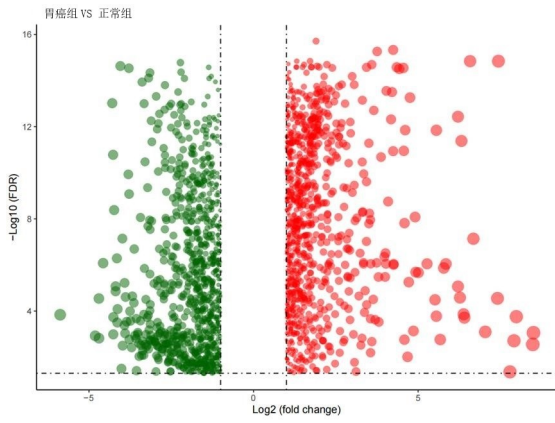
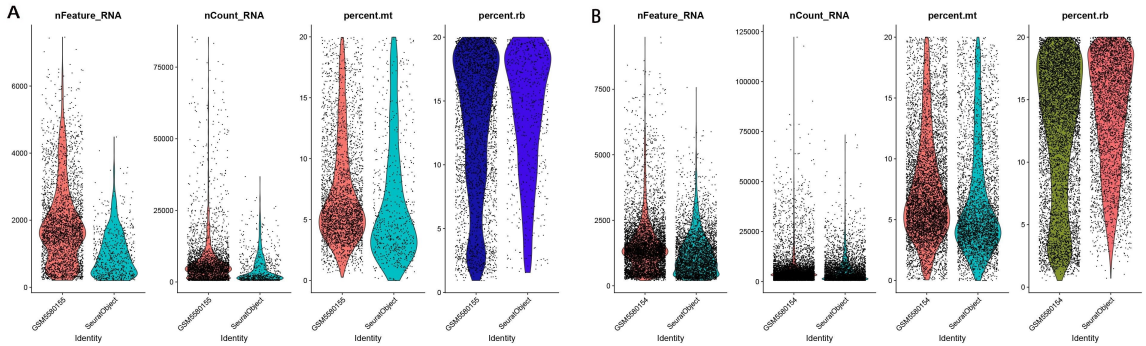


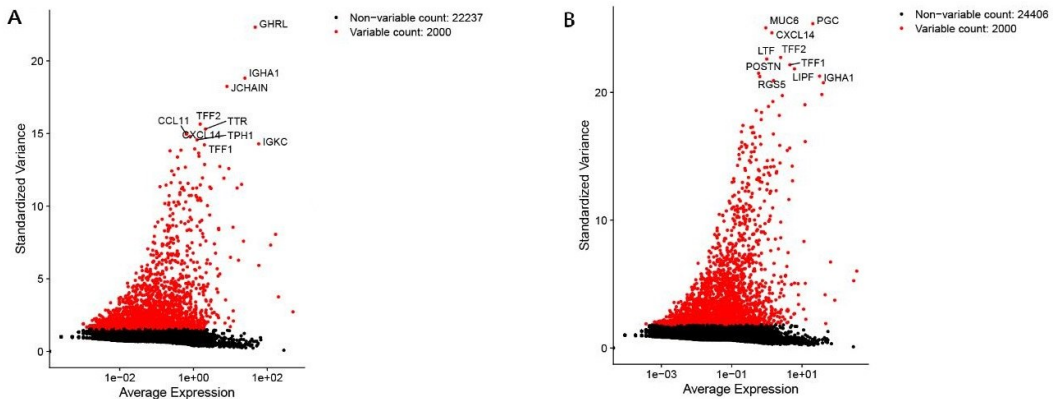
图 1 胃癌组与正常组基因差异表达分析火山图

2.2 单细胞分析结果 将 GSE167297、GSE184198 搜集得到的胃癌单细胞数据区分为正常组和胃癌组后分别进行合并,随后分别进行如下操作:对单细胞数据进行标准化(见图 2A、图 2B),并识别高变基因(见图 3A、图 3B)。对高变基因进行降维(见图 4A、图 4B),并去除批次效应,最后对单细胞进行分群聚类,两组都得到 21 个聚类(见图 5A、图 5B)。提取每个聚类中排名前 5 的基因热图(见图 6A、图 6B)对细胞进行标记并删除无法标记的细胞,最终将细胞标记为 epithelial、stromal、proliferative、T、B、myeloid(见图 7A、图 7B)。



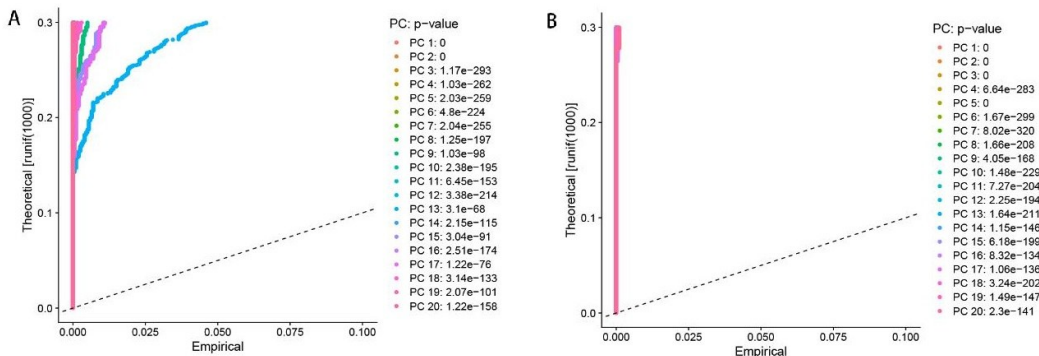
注:A. 正常组单细胞数据质控情况;B. 胃癌组单细胞数据质控情况。

图 2 单细胞数据质控相关的小提琴图



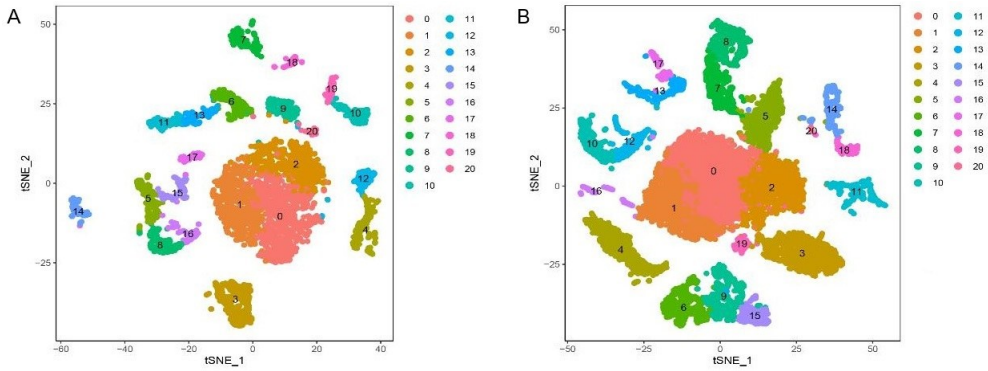
注:A. 正常组排名前 2000 的高变基因;B. 胃癌组排名前 2000 的高变基因。

图 3 鉴定高可变基因



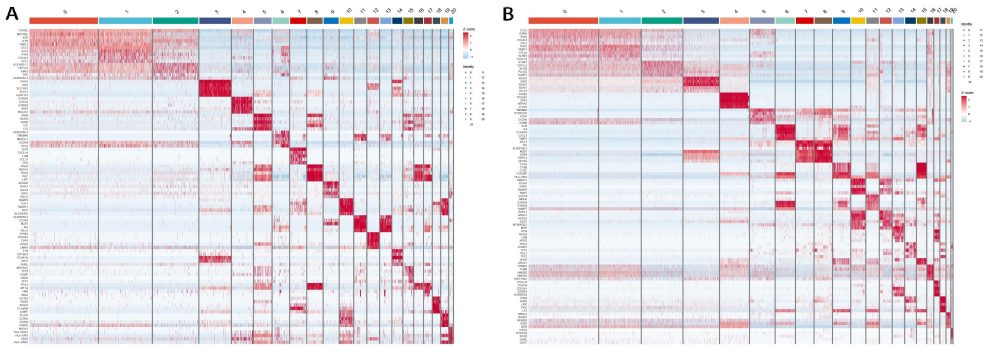
注:A. 正常组 PCA 降维;B. 胃癌组 PCA 降维。

图 4 PCA 降维



注:A. 正常组得到的 21 个细胞群聚类;B. 胃癌组得到的 21 个细胞聚类。

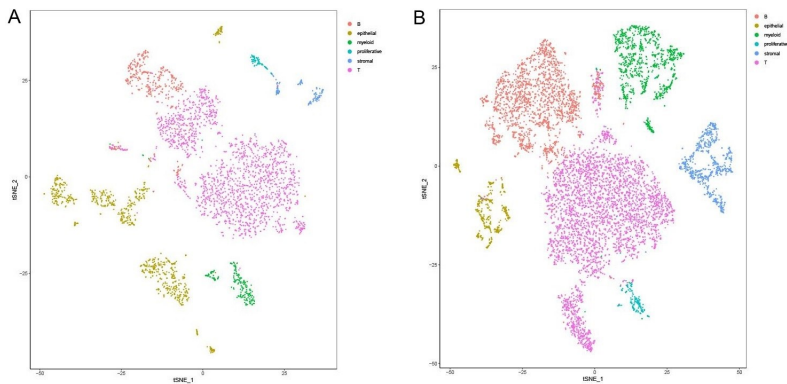
图 5 细胞聚类



注:A. 正常组中每个聚类表达量排名前 5 基因的标准分数;

B. 胃癌组中每个聚类表达量排名前 5 基因的标准分数。

图 6 每个聚类表达量排名前 5 基因的标准分数



注:A. 正常组中标记细胞类型;B. 胃癌组中标记细胞类型。

图 7 标记的细胞类型

2.3 筛选出目标基因 KRT19 单细胞正常组和胃癌组的 cluster 为 epithelial 且 avg\_log2FC 值由大到小的前 100 个基因与差异分析后的高表达基因取交集后的韦恩图(见图 8)。选择取交集后得到的 10 个胃癌的特征基因。绘制这 10 个基因的 KM 生存曲线(见图 9),筛选  $P < 0.05$  且  $HR < 1$  的基因,选择其中的 KRT19 作为本研究的目标基因。

2.4 诊断 ROC 分析结果 为了探索 KRT19 在胃癌诊断中的潜力,对其进行了 ROC 分析。ROC 曲线的 AUC 反映了模型的预测能力,分析结果显示,AUC 为 0.658,具有较高的诊断价值(见图 10)。

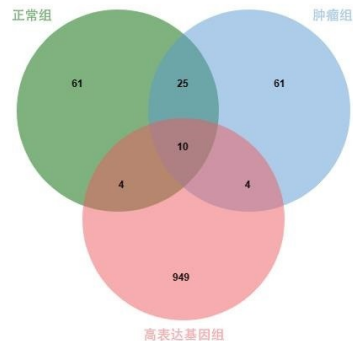
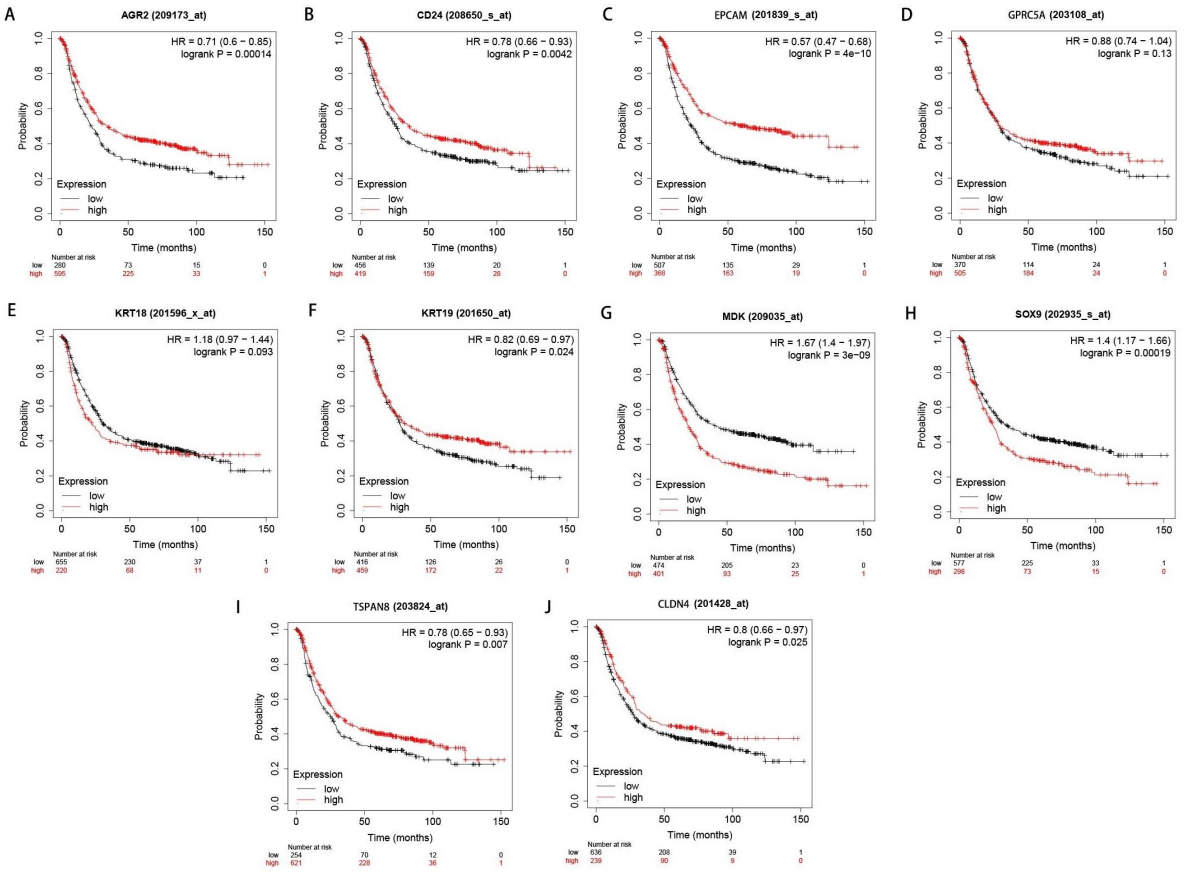


图 8 单细胞正常组和胃癌组筛选后的基因与差异分析中高表达基因的交集韦恩图



注：A. AGR2 生存曲线；B. CD24 生存曲线；C. EPCAM 生存曲线；D. GPRC5A 生存曲线；E. KRT18 生存曲线；F. KRT19 生存曲线；G. MDK 生存曲线；H. SOX9 生存曲线；I. TSPAN8 生存曲线；J. CLDN4 生存曲线。

图 9 生存曲线

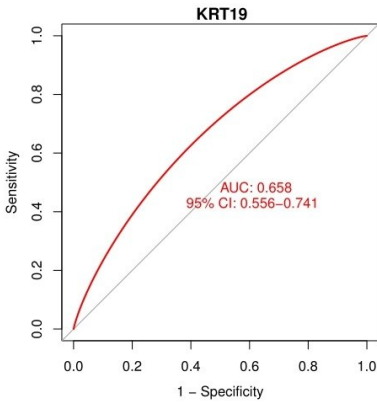
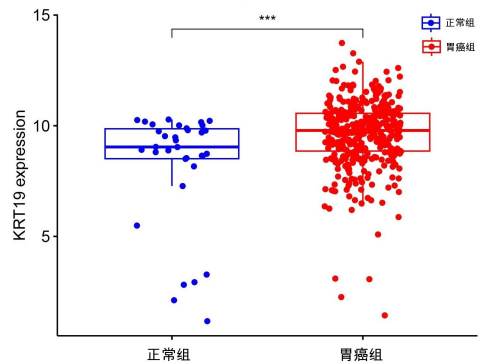


图 10 KRT19 的 ROC 诊断曲线

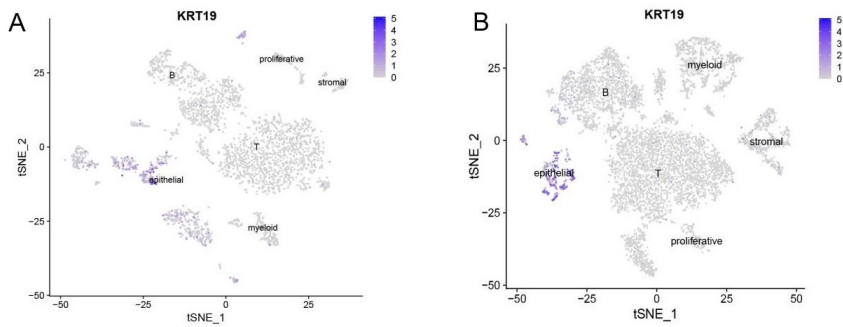
2.5 对 KRT19 进行表达量分析与单细胞分析结果  
 通过对比目标基因 KRT19 在箱型图的表达情况，得到 KRT19 在胃癌中具有高表达 ( $P < 0.001$ )，见图 11。通过对比目标基因 KRT19 在单细胞数据中的表达情况，得到目标基因在 epithelial 中有较高的表达，且在胃癌组表达更高 (见图 12A、图 12B)。通过分析目标基因 KRT19 在拟时序中的表现情况，得到目标基因 KRT19 在接近分化末期时表达程度变高 (见图 13A、图 13B)。通过对比 KRT19 高表达组 (KRT19\_high) 和 KRT19 低表达组 (KRT19\_low) 在 infercnv 的恶性

预测情况，得到 KRT19 高表达的上皮细胞可能为良性上皮细胞 (见图 14)。通过分析 KRT19 高表达组与 KRT19 低表达组的细胞通讯通路，胃癌上皮细胞在 KRT19 高低表达组之间的相互作用主要存在与 MDK (MK) 信号通路中，且可以看出主要是由在 KRT19 高表达分群产生的 (见图 15A)。通过分析 KRT19 高表达组与 KRT19 低表达组在 MK 通路的重要程度，可以看出 MK 通路主要是由 KRT19 高表达组引导 (见图 15B)。



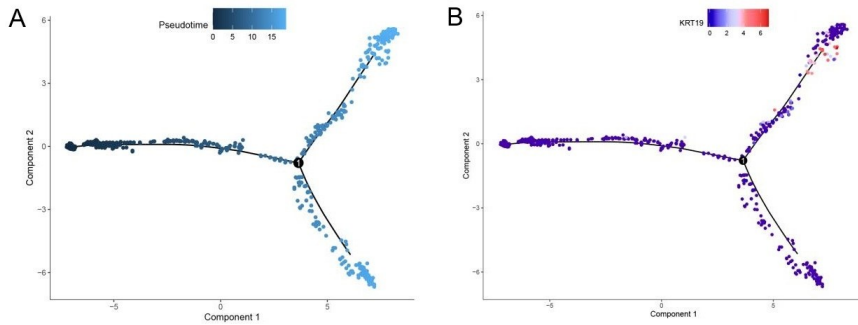
注：\*\*\*  $P < 0.001$ 。

图 11 KRT19 在正常样本与胃癌样本的差异表达情况



注:A. 正常组中 KRT19 在单细胞数据的表达情况;B. 胃癌组中 KRT19 在单细胞数据的表达情况。

图 12 KRT19 在单细胞数据的表达情况



注:A. 时序分布图;B. KRT19 在拟时序中表达情况。

图 13 拟时序图中 KRT19 的表达情况

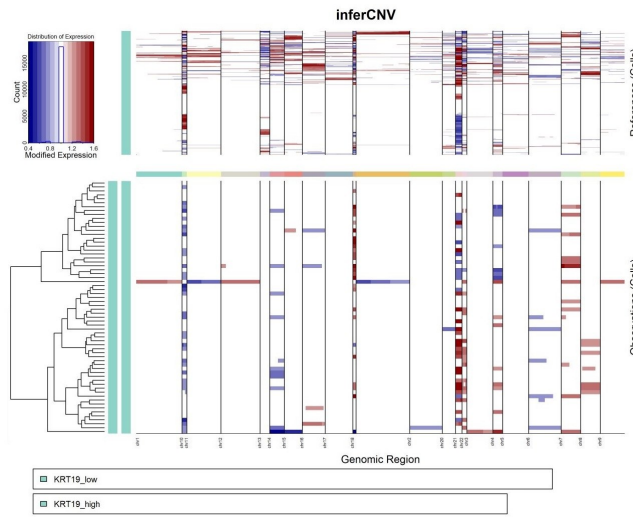
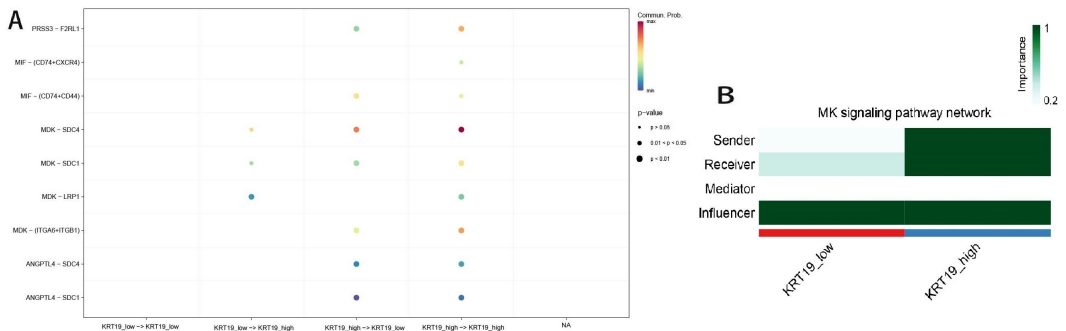


图 14 KRT19 高表达组和 KRT19 低表达组在 infercnv 恶性预测的情况



注:A. KRT19 高表达组与 KRT19 低表达组的细胞通讯通路图; B. KRT19 高表达组与 KRT19 低表达组在 MK 通路的重要程度图。

图 15 KRT19 在细胞通讯的情况

2.6 免疫浸润分析结果 对胃癌数据进行免疫浸润分析,得到各细胞在胃癌样本中的免疫浸润情况(见图 16)。分析胃癌中 KRT19 的免疫差异,得到 KRT19

的表达与 Neutrophils 浸润呈正相关(见图 17A),与 Plasma cells 浸润呈正相关(见图 17B),与 T cells CD8 浸润呈负相关(见图 17C)。

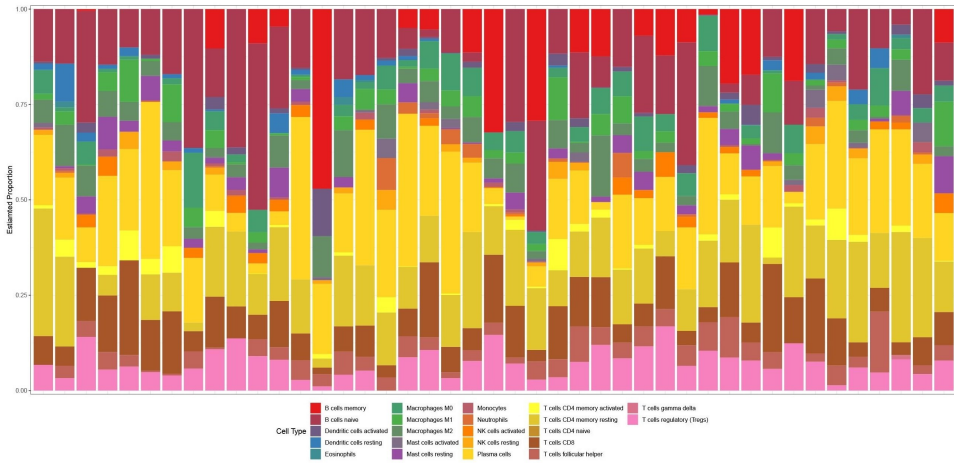
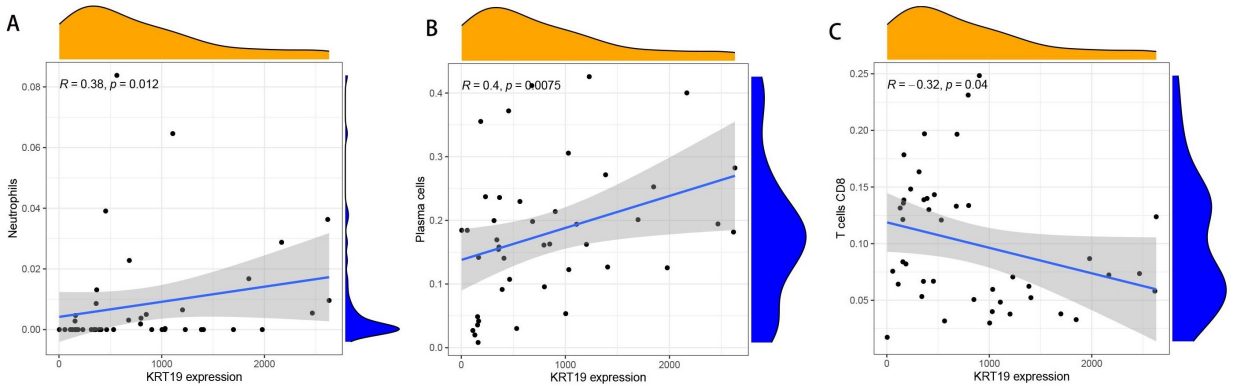


图 16 各细胞在胃癌中的免疫浸润情况



注:A. KRT19 的表达与 Neutrophils 浸润的关系;B. KRT19 的表达与 Plasma cells 浸润的关系;C. KRT19 的表达与 T cells CD8 浸润的关系。

图 17 KRT19 基因的表达与胃癌患者免疫细胞浸润的相关性

2.7 蛋白互作网络 通过 string 网站对 KRT19 进行蛋白质互作分析,得到 KRT19 与 KRT18、C17orf97、CEACAM5、AFP、EPCAM 存在互相作用关系(见图 18)。

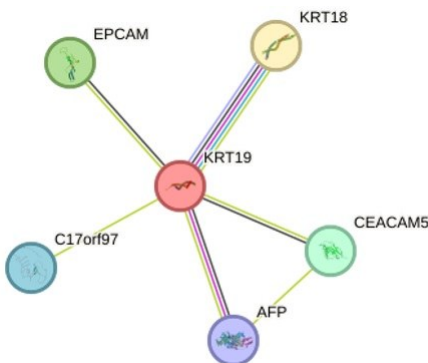
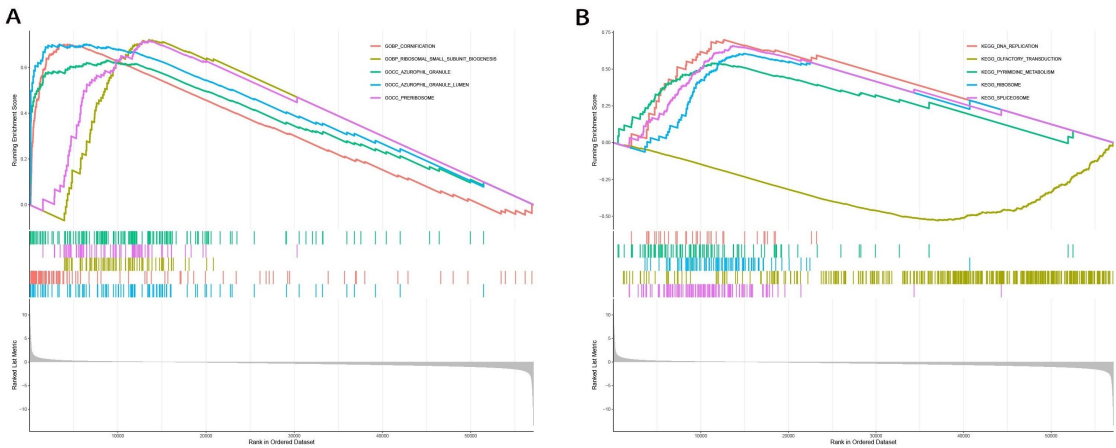


图 18 KRT19 的蛋白质互作关系

2.8 GSEA 富集分析结果 对 KRT19 进行 GSEA 富集分析。其中 GO 富集结果说明 KRT19 与角化过程、核糖体小亚基的合成过程、细胞内嗜酸性颗粒、嗜酸性颗粒腔室、预核糖体相关功能呈正相关(见图 19A)。而 KEGG 富集结果说明 KRT19 与 DNA 复制、嘧啶新陈代谢、核糖体、剪接体相关通路呈正相关但与嗅觉转导相关通路呈负相关(见图 19B)。

2.9 药物敏感性分析结果 对 KRT19 药物敏感性分析得到,在 KRT19 的高表达组中 5-氟尿嘧啶、乙二醇、BMS-345541、ERK\_6604、PD0325901、沙匹替尼的 IC50 均低于低表达组。即在 KRT19 的高表达组中 5-氟尿嘧啶、乙二醇、BMS-345541、ERK\_6604、PD0325901、沙匹替尼的药物敏感性均较低表达组高(见图 20)。



注:A. KRT19 在 GO 功能中的 GSEA 富集分析结果;B. KRT19 在 KEGG 通路中的 GSEA 富集分析结果。

图 19 KRT19 在富集分析结果

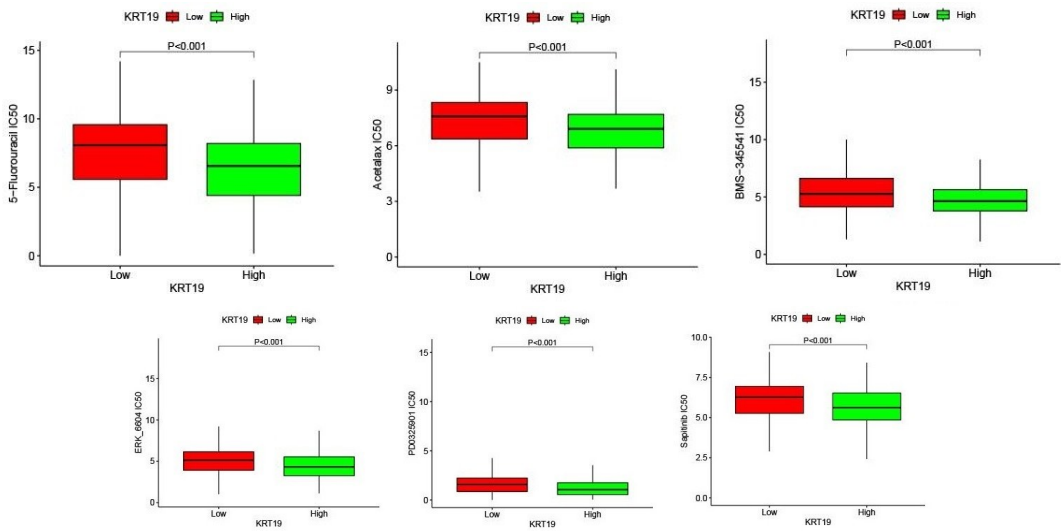


图 20 KRT19 的药物敏感性分析结果

### 3 讨论

目前,胃癌的治疗不仅限于简单的手术切除,还包括放射治疗和化疗等其他治疗方法,以控制局部疾病并提高生存率<sup>[2]</sup>。然而,早期胃癌患者通常有非特异性症状,很难被发现。因此,大多数患者在首次诊断时被诊断为晚期胃癌<sup>[3]</sup>。因此,探索胃癌的新的预后标志物可以促进对胃癌患者及时实施适当的治疗,从而提高其生存率。本研究通过单细胞测序分析结合基因差异表达分析筛选在胃癌组织中具有高表达的基因,与正常组织相比,差异具有统计学意义。随后通过 Kaplan-Meier 曲线进一步判断可以作为胃癌新预后标志物的基因,然而已有研究证明 CD24 在胃癌中有预测作用<sup>[14]</sup>,EPCAM 被证明在胃癌中有预测作用<sup>[14]</sup>,CLDN4 被证明可以减弱胃癌细胞增殖<sup>[15]</sup>,TSPAN8 被证明可以促进胃癌的生长<sup>[16]</sup>,AGR2 被证明与胃癌的进展有关<sup>[17]</sup>。最后选择 KRT19 作为本研究的目标基因。

KRT 是构成细胞骨架的中间丝蛋白,通过与一系

列细胞蛋白(包括激酶-受体-衔接子和效应子)的动态相互作用来调节细胞周期、细胞凋亡、细胞对应激和死亡的反应<sup>[18]</sup>。有 20 种类型的 KRT 按分子量和等电点分类。这些 KRT 分为 2 种亚型:低分子量酸性 I 型和高分子量碱性或中性 II 型<sup>[19]</sup>。KRT19 是已知最小的(40 kDa) I 型 KRTs,与某些疾病有关<sup>[20]</sup>。目前,KRT19 被鉴定为多种癌症的肿瘤相关生物标志物,包括食管鳞状细胞癌<sup>[21]</sup>、宫颈癌<sup>[22]</sup>、乳腺癌<sup>[23]</sup>和肝细胞癌<sup>[24]</sup>。

然而,胃癌是第五大常见癌症,也是癌症相关死亡的第三大常见原因,占全球近 800 000 人死亡<sup>[25]</sup>。为了研究 KRT19 对于胃癌的预后价值,本研究使用 TCGA 数据库的 TCGA-胃癌转录组数据进行差异分析,差异分析结果显示 KRT19 在胃癌中具有高表达。为了进一步研究 KRT19 在细胞之间的情况,本研究使用 GEO 数据库的胃癌单细胞数据进行聚类 and 注释后,发现 KRT19 在胃癌样本和正常样本上皮细胞均有较高表达,且在胃癌样本的表达变高。随后进行拟时序分

析,发现正常组中 KRT19 并不是拟时序差异基因,但在胃癌样本中却是作为拟时序差异基因,在分化末端的上皮细胞群具有较强的表达。为了探究 KRT19 所在的上皮细胞群,是否为恶性上皮细胞,随即提取胃癌样本中的上皮细胞群进行再次聚类,并根据 KRT19 的高低表达进行注释,而后进行 infercnv 分析,结果显示 KRT19 高表达的上皮细胞是良性上皮细胞。所以认为 KRT19 的高表达可能表示胃癌良性上皮细胞的出现。

随后为了探究 KRT19 所在的上皮细胞群是哪条通路发挥的作用,本课题组使用了 Cellchat 进行了细胞通讯分析其结果,KRT19 高表达促进上皮细胞细胞信号通路的之间的交流。而在众多信号通路中,主要是 MDK(MK)信号通路发挥着影响 KRT19 所在胃癌上皮细胞间的细胞间信息交流。而在这个 MK 信号通路中有 3 个配体-受体对(MDK-SDC4、MDK-SDC1、MDK-LRP1)对其进行贡献。MDK(MK)是一种可溶性蛋白,在癌症等各种疾病中常常显著升高,因此它可以作为一种有价值的疾病标志物<sup>[26]</sup>,在许多类型的癌症中,MDK(MK)已被证明过度表达<sup>[26]</sup>,特别是在肿瘤进展到更晚期时<sup>[27]</sup>。MDK(MK)目前已被证实正常和疾病状态下的上皮组织中的作用,包括在上皮细胞增殖、迁移、上皮-间质转化等方面的功能<sup>[28]</sup>。所以本研究认为之所以是 MK 信号通路在 KRT19 表达的上皮细胞间的信息交流中发挥着主要影响作用,是可能与 KRT19 高表达促进良性上皮细胞的增殖有关,结合之前的拟时序分析,KRT19 高表达的上皮细胞群可能是在分化末期出现,这有可能是促进 MDK 在肿瘤进展到更晚期过度表达的因素之一。

本研究还用 TCGA 数据进行免疫浸润分析,发现 KRT19 的表达与 Neutrophils 浸润呈正相关,与 Plasma cells 浸润呈正相关关系,与 T cells CD8 浸润呈负相关。结合 KRT19 的 KM 生存曲线说明 KRT19 在胃癌中可能通过调控 Neutrophils 和 Plasma cells 的表达从而抑制胃癌的进展。然后结合之前的拟时序分析,本研究认为可能是因为 KRT19 表达变强的时候,T 细胞已经进行了分化,故与 T cells CD8 浸润呈负相关,同时 B 细胞浸润程度增加可能是 T 细胞分化产生的结果。通过蛋白互作网络,得到 KRT19 与 KRT18、C17orf97、CEACAM5、AFP、EPCAM 存在互相作用关系。其中 KRT18 选择性调控细胞增殖和凋亡基因的表达<sup>[29]</sup>。EPCAM 是一种 40 kDa 的 I 型跨膜糖蛋白,已知在上皮癌中高表达,可作为预后因素<sup>[30]</sup>。string 数据库说明了 EPCAM 可作黏膜上皮处肠上皮细胞和上皮内淋巴细胞之间的物理同性相互作用分子,作为抵抗黏膜感染的第一道防线提供免疫屏障。所以结合此前的分析 KRT19 可能与 KRT18 相互作用调控良

性上皮细胞的产生与增殖。而 EPCAM 具有使黏膜上皮处肠上皮细胞和上皮内淋巴细胞之间的物理同性相互作用的功能,联系到 KRT19 与 Plasma cells 和 Neutrophils 的浸润的正相关性还有胃癌组的拟时序结果,本研究认为在分化上皮细胞后,上皮细胞内的高表达 KRT19 可能与同样处在胃癌上皮细胞的高表达 EPCAM 相互作用,引起或促进了上皮细胞和上皮内淋巴细胞之间的物理同性的相互作用,进而产生了随着 KRT19 的表达增加,Neutrophils 的浸润和 Plasma cells 的浸润程度加深。随后进行的 GSEA 富集分析中 GO 富集结果说明 KRT19 与角化过程、核糖体小亚基的合成过程、细胞内嗜酸性颗粒、嗜酸性颗粒腔室、预核糖体相关功能呈正相关。而 KEGG 富集结果说明 KRT19 与 DNA 复制、嘧啶新陈代谢、核糖体、剪接体相关通路呈正相关但与嗅觉转导相关通路呈负相关。目前已知在蛋白质合成过程中,起始 tRNA 在小核糖体亚基(SSU)上进行装配,以识别 mRNA 的起始密码子<sup>[31]</sup>。而 KRT19 与核糖体小亚基的合成过程正相关说明了 KRT19 的高表达可能促进 tRNA 在 SSU 上的装配从而推动蛋白质的合成。目前已知,在细胞核内,核糖体蛋白质-转运蛋白复合物与预核糖体特定区域上的其他组分相互作用,从而促进核糖体蛋白质的同步组装<sup>[32]</sup>。而 KRT19 与预核糖体相关功能呈正相关说明了 KRT19 的高表达可能有促进核糖体蛋白质的同步组装。结合之前单细胞结果讨论,课题组认为 KRT19 的高表达可能通过促进蛋白质的合成与核糖体蛋白质的同步组装从而推动胃癌良性上皮细胞的出现和增殖。最后,药物敏感性分析结果表明,药物对 KRT19 的高表达具有更高的敏感性,可能是因为 KRT19 在胃癌中有更高的表达,所以药物对其高表达的敏感性更高。

综上所述,KRT19 可能在分化末期的胃癌上皮细胞中具有更高的表达,可能具有促进胃组织的良性上皮细胞出现和增殖的作用,对胃癌的发展具有良好预后的作用。但应注意本研究研究的主要局限性,结果基于生物信息学分析,未来需要进行进一步的实验验证,以验证胃癌中 KRT19 表达的生物学意义。

#### 参考文献:

- [1] SUNG H, FERLAY J, SIEGEL R L, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA Cancer J Clin, 2021, 71(3): 209-249.
- [2] CHARALAMPAKIS N, TSAKATIKAS S, SCHIZAS D, et al. Trimodality treatment in gastric and gastroesophageal junction cancers: current approach and future perspectives[J]. World J Gastrointest Oncol, 2022, 14(1): 181-202.
- [3] NIZIOI M, ZINCZUK J, ZAREBA K, et al. Increased ten-

- sin 4 expression is related to the histological type of gastric cancer[J]. *World J Clin Oncol*, 2021, 12(12): 1202-1214.
- [4] FRADETTE J, GERMAIN L, SESHIAIAH P, et al. The type I keratin 19 possesses distinct and context-dependent assembly properties [J]. *J Biol Chem*, 1998, 273(52): 35176-35184.
- [5] COULOMBE P A, WONG P. Cytoplasmic intermediate filaments revealed as dynamic and multipurpose scaffolds [J]. *Nat Cell Biol*, 2004, 6(8): 699-706.
- [6] HENDRIX M J C, SEFTOR E A, CHU Y W, et al. Role of intermediate filaments in migration, invasion and metastasis[J]. *Cancer Metastasis Rev*, 1996, 15(4): 507-525.
- [7] GIRES O, PAN M, SCHINKE H, et al. Expression and function of epithelial cell adhesion molecule EpCAM: where are we after 40 years? [J]. *Cancer Metastasis Rev*, 2020, 39(3): 969-987.
- [8] NEESSE A, GRIESMANN H, GRESS T M, et al. Claudin-4 as therapeutic target in cancer[J]. *Arch Biochem Biophys*, 2012, 524(1): 64-70.
- [9] FUJIWARA-TANI R, MORI S, OGATA R, et al. Claudin-4: a new molecular target for epithelial cancer therapy [J]. *Int J Mol Sci*, 2023, 24(6): 5494.
- [10] BAGÓ J R, ALIEVA M, SOLER C, et al. Endothelial differentiation of adipose tissue-derived mesenchymal stromal cells in glioma tumors: implications for cell-based therapy[J]. *Mol Ther*, 2013, 21(9): 1758-1766.
- [11] PAUNESKU T, MITTAL S, PROTIC M, et al. Proliferating cell nuclear antigen (PCNA): ringmaster of the genome [J]. *Int J Radiat Biol*, 2001, 77(10): 1007-1021.
- [12] MASON D Y, CORDELL J L, BROWN M H, et al. CD79a: a novel marker for B-cell neoplasms in routinely processed tissue samples[J]. *Blood*, 1995, 86(4): 1453-1459.
- [13] EDWARDS D K V, WATANABE-SMITH K, ROFELTY A, et al. CSF1R inhibitors exhibit antitumor activity in acute myeloid leukemia by blocking paracrine signals from support cells[J]. *Blood*, 2019, 133(6): 588-599.
- [14] GÓMEZ-GALLEGOSÁ A, RAMÍREZ-VIDAL L, BECERRIL-RICO J, et al. CD24 + CD44 + CD54 + EpCAM + gastric cancer stem cells predict tumor progression and metastasis: clinical and experimental evidence[J]. *Stem Cell Res Ther*, 2023, 14(1): 16.
- [15] LUO J, WANG H M, CHEN H J, et al. CLDN4 silencing promotes proliferation and reduces chemotherapy sensitivity of gastric cancer cells through activation of the PI3K/Akt signalling pathway[J]. *Exp Physiol*, 2020, 105(6): 979-988.
- [16] WEI L S, LI Y, SUO Z M. TSPAN8 promotes gastric cancer growth and metastasis via ERK MAPK pathway [J]. *Int J Clin Exp Med*, 2015, 8(6): 8599-8607.
- [17] ZHANG J, JIN Y M, XU S N, et al. AGR2 is associated with gastric cancer progression and poor survival [J]. *Oncol Lett*, 2016, 11(3): 2075-2083.
- [18] COULOMBE P A, WONG P. Cytoplasmic intermediate filaments revealed as dynamic and multipurpose scaffolds[J]. *Nat Cell Biol*, 2004, 6(8): 699-706.
- [19] RHEE H, KIM H Y, CHOI J H, et al. Keratin 19 expression in hepatocellular carcinoma is regulated by fibroblast-derived HGF via a MET-ERK1/2-AP1 and SP1 axis[J]. *Cancer Res*, 2018, 78(7): 1619-1631.
- [20] WU Y J, RHEINWALD J G. A new small (40 kd) keratin filament protein made by some cultured human squamous cell carcinomas[J]. *Cell*, 1981, 25(3): 627-635.
- [21] ZHANG J T, WANG K J, ZHANG J Z, et al. Using proteomic approach to identify tumor-associated proteins as biomarkers in human esophageal squamous cell carcinoma[J]. *J Proteome Res*, 2011, 10(6): 2863-2872.
- [22] OKAMOTO S, NIIKURA H, NAKABAYASHI K, et al. Detection of sentinel lymph node metastases in cervical cancer: assessment of KRT19 mRNA in the one-step nucleic acid amplification (OSNA) method[J]. *Gynecol Oncol*, 2013, 130(3): 530-536.
- [23] JU J H, YANG W, LEE K M, et al. Regulation of cell proliferation and migration by keratin19-induced nuclear import of early growth response-1 in breast cancer cells [J]. *Clin Cancer Res*, 2013, 19(16): 4335-4346.
- [24] ZENG S S, YAMASHITA T, KONDO M, et al. The transcription factor SALL4 regulates stemness of EpCAM-positive hepatocellular carcinoma[J]. *J Hepatol*, 2014, 60(1): 127-134.
- [25] SMYTH E C, NILSSON M, GRABSCH H I, et al. Gastric cancer[J]. *Lancet*, 2020, 396(10251): 635-648.
- [26] JONES D R. Measuring midkine: the utility of midkine as a biomarker in cancer and other diseases [J]. *Br J Pharmacol*, 2014, 171(12): 2925-2939.
- [27] JONO H, ANDO Y. Midkine: a novel prognostic biomarker for cancer[J]. *Cancers*, 2010, 2(2): 624-641.
- [28] MONMA F, HOZUMI Y, IKEMATSU S, et al. Expression of midkine in normal human skin, dermatitis and neoplasms: association with differentiation of keratinocytes [J]. *J Dermatol*, 2013, 40(12): 980-986.
- [29] CHEN B, XU X M, LIN D D, et al. KRT18 modulates alternative splicing of genes involved in proliferation and apoptosis processes in both gastric cancer cells and clinical samples[J]. *Front Genet*, 2021, 12: 635429.
- [30] YONAIYAMA S, TOYOKI Y, MOROHASHI S, et al. Epithelial cell adhesion molecule (EpCAM) overexpression is correlated with malignant potentials of intraductal papillary mucinous neoplasms (IPMNs) of the pancreas [J]. *Biomed Res*, 2013, 34(2): 87-95.
- [31] KHAWAJA A, ITOH Y, REMES C, et al. Distinct pre-initiation steps in human mitochondrial translation[J]. *Nat Commun*, 2020, 11(1): 2932.
- [32] KRESSLER D, BANGE G, OGAWA Y, et al. Synchronizing nuclear import of ribosomal proteins with ribosome assembly[J]. *Science*, 2012, 338(6107): 666-671.