

COL10A1在胃癌中的作用机制及其临床意义

陈 融¹, 姜相君^{2*}

¹青岛大学医学部, 山东 青岛

²青岛市市立医院消化内科, 山东 青岛

收稿日期: 2023年5月28日; 录用日期: 2023年6月23日; 发布日期: 2023年6月30日

摘要

目的: 胃癌(gastric cancer, GC)作为最常见的消化道恶性肿瘤, 由于早期症状不明显, 发现时多为晚期, 预后较差。COL10A1过去主要研究其在骨关节疾病中的作用, 而近年来更多地关注于该基因在肿瘤中的作用, 其可能通过改变肿瘤微环境的成分, 发挥促进肿瘤发生、转移等方面的能力, 具有成为多种肿瘤诊断标志物或治疗相关靶点的潜力, 而其在胃癌中作用的研究较少, 本研究目的在于利用生物信息学分析的方法, 探究COL10A1在对于胃癌发生、发展中的作用以及在肿瘤诊断、预后预测及胃癌治疗等方面的临床应用价值。方法: 收集TCGA数据库中胃癌患者的相关数据, 联合GTEx等其他数据库数据, 使用R软件中相关R包分析COL10A1在胃癌与普通组织中的表达差异。进而分析COL10A1高低表达组之间的基因表达情况得到差异表达基因, 对这些差异表达基因进行功能富集分析(GO, KEGG, GSEA), 探究目标基因COL10A1差异表达可能影响的细胞功能、通路。使用Estimate算法对COL10A1高低表达组样本进行打分, 然后比较高低表达组之间评分, 评估COL10A1的表达差异对细胞外基质的影响。同时使用CIBERSORT算法对COL10A1高低表达组中样本的免疫细胞浸润进行分析, 比较高低表达组之间免疫细胞浸润情况的不同, 分析COL10A1对于肿瘤组织中的免疫细胞浸润的影响。为探究COL10A1在临床中的作用, 首先根据TCGA数据库中临床数据进行生存分析, 分析COL10A1对于总生存期、无病生存期的影响, 进而对COL10A1表达量及其他常见因素(如年龄、性别、TNM分期、Stage分期等)进行单因素COX分析与多因素COX分析, 并构建多因素COX风险比例模型。根据TCIA数据库中数据比较目标基因高低表达组之间免疫治疗效果, 并评估COL10A1与常见免疫检查点表达之间表达相关性, 探究其与免疫治疗的相关性。通过上述方法探究COL10A1在胃癌中可能参与的机制、对于预后、免疫治疗效果的影响。结果: 通过分析TCGA、GTEx数据库数据, 目标基因COL10A1在包括胃癌在内的多种肿瘤中显著高表达, 结合所收集样本测序结果, 目标基因在胃癌组织与正常胃组织中表达差异明显。通过分析数据库中数据表明COL10A1表达量与患者预后相关, 对于总生存期, 高表达胃癌患者的总生存期较短, 预后较差。经单因素及多因素COX分析, COL10A1表达量是患者预后的独立危险因素, 并根据TCGA数据库数据, 构建了预后预测模型。在细胞外基质的组成方面发挥作用, 同时可能参与了TGF-β通路及PI3K-Akt信号通路, 两种通路均与肿瘤的发生发展相关。通过进行Estimated、CIBERSORT分析, COL10A1高低表达组患者细胞外基质差异明显, 高表达组患者肿瘤纯度更低、免疫细胞浸润数量更多, 于是其表达与肿瘤微环境中免疫细胞浸润呈正相关。高表达组免疫治疗效果较差, 可能与目标基因和常见免疫检查点表达呈正相关有关。预示着该基因有着作为治疗靶点的潜力。结论: COL10A1在胃癌中高表达, 当COL10A1表达上调时胃癌患者的总体生存期显著缩短, 但对无进展生存期无明显影响, 提示其与患者的预后不良有关。COL10A1与肿瘤微环境的构成有关, 与免疫细胞浸润和免疫检查点的表达存在显著性相关。通过分析TCIA数据库中数据, COL10A1高表达会使部分患者的免疫治疗效果差, 提示COL10A1可能是一个潜在的免疫治疗靶点。

*通讯作者。

关键词

胃癌，基因表达数据库，癌症基因组图谱，X型胶原 α 1

The Mechanism and Clinical Significance of COL10A1 in Gastric Cancer

Rong Chen¹, Xiangjun Jiang^{2*}

¹Medical College, Qingdao University, Qingdao Shandong

²Department of Gastroenterology, Qingdao Municipal Hospital, Qingdao Shandong

Received: May 28th, 2023; accepted: Jun. 23rd, 2023; published: Jun. 30th, 2023

Abstract

Objective: Gastric cancer (GC), as the most common malignant tumor of the digestive tract, is often found in the late stage due to its unclear early symptoms and poor prognosis. COL10A1 used to mainly study its role in bone and joint diseases, but in recent years more attention has been paid to the role of this gene in tumors. It may play a role in promoting tumor occurrence, metastasis and other aspects by changing the composition of the tumor microenvironment, and has the potential to become a variety of tumor diagnostic markers or treatment related targets, while there is less research on its role in gastric cancer. The purpose of this study is to use bioinformatics analysis methods to explore the role of COL10A1 in the occurrence and development of gastric cancer, as well as its clinical application value in tumor diagnosis, prognosis prediction, and treatment of gastric cancer.

Method: Collect relevant data of gastric cancer patients in the TCGA database, combined with data from other databases such as GTEx, and analyze the expression difference of COL10A1 between gastric cancer and common tissues using the relevant R package in R software. Further analysis of the gene expression between the high and low expression groups of COL10A1 was conducted to identify differentially expressed genes. Functional enrichment analysis (GO, KEGG, GSEA) was performed on these differentially expressed genes to explore the potential cellular functions and pathways affected by the differential expression of the target gene COL10A1; use the Estimate algorithm to score the samples of COL10A1 high and low expression groups, and then compare the scores between the high and low expression groups to evaluate the impact of COL10A1 expression differences on the extracellular matrix. Simultaneously, the CIBERSORT algorithm was used to analyze the immune cell infiltration of samples in the high and low expression groups of COL10A1, compare the differences in immune cell infiltration between the high and low expression groups, and analyze the impact of COL10A1 on immune cell infiltration in tumor tissue. To explore the role of COL10A1 in clinical practice, survival analysis was first conducted based on clinical data from the TCGA database, analyzing the impact of COL10A1 on overall survival and disease-free survival. Then, single factor COX analysis and multi factor COX analysis were conducted on the expression level of COL10A1 and other common factors (such as age, gender, TNM stage, Stage stage, etc.), and a multi factor COX risk proportion model was constructed. To analyze the expression correlation between COL10A1 and common immune checkpoint expression, to delve into its relevance to immunotherapy, and to compare the effect of immunotherapy between groups of high and low expression of target genes using data from the TCIA database, the drug sensitivity of high and low expression groups to some chemotherapeutic agents was analyzed using pRRophetic. Through the above methods, we investigated the possible mechanisms of COL10A1 involvement in GC and its impact on the prognosis and immunotherapeutic effect.

Result: By analyzing the TCGA and GTEx database

data, the target gene COL10A1 was significantly over expressed in various tumors, including gastric cancer. Combined with the sequencing results of the collected samples, the target gene showed significant differences in expression between gastric cancer tissue and normal gastric tissue. By analyzing the data in the database, it was found that the expression level of COL10A1 is related to patient prognosis. For overall survival, high expression gastric cancer patients have a shorter overall survival period and a poorer prognosis. After univariate and multivariate COX analysis, the expression of COL10A1 was found to be an independent risk factor for patient prognosis. Based on the TCGA database data, a prognosis prediction model was constructed. Analysis by GO, KEGG, and GSEA concluded that COL10A1 plays a role in the composition of the extracellular matrix and may be involved in the TGF- β pathway and PI3K-Akt signaling pathway, both of which are usually risk factors for tumor development according to the current findings. Through the estimated and CIBERSORT analysis, the extracellular matrix of patients with high and low expression of COL10A1 was significantly different. The tumor purity of patients with high expression of COL10A1 was lower, and the number of immune cells infiltrated was more. Therefore, its expression was positively correlated with the infiltration of immune cells in the tumor microenvironment. The immunotherapy effect of the high expression group is poor, which may be positively correlated with the expression of target genes and common immune checkpoints, indicating the potential of this gene as a therapeutic target. Conclusion: COL10A1 is highly expressed in gastric cancer. When COL10A1 expression is upregulated, the overall survival period of gastric cancer patients is significantly shortened, but it has no significant impact on progression free survival, indicating that it is related to poor prognosis of patients. COL10A1 is related to the composition of tumor microenvironment, and is significantly related to the infiltration of immune cells and the expression of immune checkpoints. By analyzing the data in the TCIA database, it was found that high expression of COL10A1 can lead to poor immunotherapy efficacy in some patients, suggesting that COL10A1 may be a potential immunotherapy target.

Keywords

Gastric Cancer, Gene Expression Database, Cancer Genome Map, X-Type Collagen α One

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

胃癌作为全球性的健康问题，作为最为常见的消化道恶性肿瘤之一，在 2020 年全球新增病例超过 100 万例，发病率排名第五，死亡率排名第四，其中东亚地区的发病率显著高于其他地区[1]。近年来胃镜检查的推广，使得胃癌的早诊早治率较前有明显提升，我国胃癌的发病率较前明显降低[2]。但尽管在内镜诊治、手术治疗、化学药物治疗、靶向药物治疗等方面有了很大进步，但当前胃癌患者的 5 年生存率仍较低[3]，在我国，胃癌的发病率和因胃癌死亡的人数仍位于各种肿瘤的前列。并且因为胃癌患者在肿瘤早期并没有显著的症状，当因为上消化道出血、胃部不适或幽门梗阻等症状就诊时，患者往往已处于胃癌进展期，部分患者已经失去手术根治机会。目前对于不能接受手术的患者进行抗肿瘤治疗时首选铂类及氟尿嘧啶联合序贯化疗方案进行抗肿瘤治疗[4]；对于特定靶点阳性的患者，可以同时联合一些分子靶向药物，如曲妥珠单抗、雷莫芦单抗等进行治疗。根据上一次欧洲医学肿瘤学会(ESMO) 2020 年虚拟会议上发表的 CheckMate 649 研究结果，在晚期 GC 中，化疗和免疫疗法的结合似乎是最有前途的方法[5]，但即使是经过上述药物联合治疗，胃癌患者的总体中位生存时间仍然是难以超过 12 个月[3] [6]。近几年来免疫检查点抑制剂(immune checkpoint inhibitors, ICIs)应用为主的免疫治疗为胃癌患者带来了新

的希望。目前已有的 ICIs 如派姆单抗、纳武单抗等在临床试验中表现出了较好的治疗效果[7] [8]，ICIs 的应用在中国临床肿瘤医学会(CSCO)的指南中的地位越来越高，覆盖范围也越来越广，目前 ICIs 的应用已从三线治疗升级为一线治疗。但尽管当前晚期胃癌行 ICIs 联合化疗的客观缓解率(overall response rate, ORR)较高，仍有一部分患者效果欠佳，且治疗相关不良事件(treatment-related adverse event, TRAE)发生率较高。目前许多研究表明，不同患者肿瘤中免疫细胞浸润程度以及免疫检查点表达的水平差异可能是造成患者对于免疫治疗耐药性和不良反应的原因。因此，探究与肿瘤免疫细胞浸润与肿瘤细胞免疫检查点表达的新的治疗靶点以及准确的免疫浸润预后标志物，是提高胃癌患者免疫治疗效率所急需的。

X α 1 型胶原蛋白(Collagen type X alpha 1, COL10A1)是属于胶原蛋白家族的一员，该基因编码的是参与软骨内成骨过程的一种小链状胶原蛋白[9]，过去该蛋白主要认为与软骨发育不良引起的 Schmid 型干骺端软骨发育异常(Schmid type metaphyseal chondrodysplasia, SMCD)以及骨关节炎有关[10] [11]，近期研究发现该蛋白是细胞外基质中的一种主要基质成分。根据当前研究，细胞外基质已被确定在肿瘤细胞的生长、分化、凋亡、转移等各个方面发挥重要作用[12]。近几年的研究显示，COL10A1 在各种实体肿瘤中的高表达可能与肿瘤的血管生成有关[13]，而且 COL10A1 在结肠癌[13] [14]、乳腺癌[15] [16]、肺癌[17]等多种肿瘤中与患者的预后有关，是影响总生存率的独立威胁因素。但当前对于 COL10A1 在胃癌的作用尚不明确。

肿瘤微环境(tumor microenvironment, TME)指的是肿瘤的发生、生长及转移等过程中肿瘤细胞所处的内外环境，它不仅包括肿瘤所在的组织结构、功能和代谢，而且于肿瘤细胞自身的内在环境有关。肿瘤微环境的组成不仅仅与肿瘤细胞自身有关，同时与免疫细胞、基质细胞、内皮细胞以及细胞外基质分子等多种成分有关，他们在调节肿瘤发生、发展的过程中以及抗肿瘤治疗的效果等多种方面发挥着关键的作用[18]。依托于基因测序技术的发展，目前生物信息学分析已经用于众多肿瘤的研究之中，近年来肿瘤微环境的研究是当前肿瘤研究的热点之一，使用生物信息学分析来研究肿瘤微环境也已在多种肿瘤中得到应用，如乳腺癌[19] [20]、结直肠癌[21]、胆囊癌[22] [23]、肾细胞癌[24]等肿瘤的研究中。为明确 COL10A1 是否影响了肿瘤微环境的构成，本研究通过使用 CIBERSORT 研究目标基因差异表达与肿瘤微环境之间的相关性，通过比较高低表达组之间免疫细胞浸润情况，来比较两者之间肿瘤微环境差异；同时还可以通过使用 Estimate 算法来分别评价目标基因高低表达组之间得分，来比较两者肿瘤微环境的差异[25]：Estimate 算法不仅将免疫细胞的数量纳入评价标准当中，同时也将基质细胞以及肿瘤细胞纯度等因素纳入对肿瘤微环境的评价之中。通过两种方法，来比较目标基因高低表达对于肿瘤微环境构成的影响。

本研究通过分析公共数据库 TCGA、GTEx 等数据库中数据以及青岛市市立医院行手术治疗的胃癌患者的肿瘤组织测序数据，探究了 COL10A1 在胃癌中可能参与的进程，COL10A1 高表达与患者预后的相关性以及 COL10A1 与免疫细胞浸润相关的关系。

2. 资料和方法

2.1. 数据下载及处理

本研究中所用到的数据主要是由 TCGA 中的 STAD 项目中的可公开使用的转录组数据，该数据由 GDC 官网下载，下载时将数据类型选择为“transcriptomeprofiling”，数据种类选为“Gene Expression Quantification”，使用 GDC 官方下载工具下载 RNA-Seq 测序数据，共 407 例，其中测序数据中共有胃癌标本 375 例，正常癌旁组织 32 例。然后下载对应的临床数据，同样在官网将数据类型选择为“Clinical”，数据种类选为“Clinical Supplement”，同样通过 GDC 官网下载工具下载。下载完成后进行数据的初步处理，对于基因表达量部分，首先将 RNA-Seq 数据由 FPKM (Fragments Per Kilobase per Million)格式转换为 TPM (transcripts per million reads)格式，然后根据 TPM 值将所有展示的基因表达量转换为经 log2

(TPM + 1)转换后的数值；对于临床数据部分，将所下载数据中所需要用到的部分进行提取、合并，包括样本对应患者的基础信息，包括性别、年龄、生存状态等，以及疾病相关数据，包括肿瘤分期、生存时间等。GTEX 中正常胃组织测序数据由 UCSCXENA 官网下载，该数据是经过 Toil 流程统一处理过的 TCGA 项目中的测序数据以及 GTEX 中的正常组织测序数据，已经过 TPM 转化，用于表达差异分析。

2.1.1. 差异表达分析

表达差异分析部分主要分为两个部分进行，一个是通过使用 TIMER2.0 在线工具，对 COL10A1 在多种肿瘤及肿瘤旁正常组织的表达量进行对比，大致了解 COL10A1 在各种肿瘤中的表达情况。作为验证，同时将使用 R 软件分析合并了 GTEX 与 TCGA 的数据，从而得出 COL10A1 在各肿瘤及正常组织中的表达差异。

然后具体到胃癌中，使用 R 软件中的 limma 包来比较正常癌旁组织与胃癌组织中的 COL10A1 表达量的不同，然后使用 ggpubr 包将结果进行可视化转换，得到差异表达图形。由于 TCGA 中的 STAD 项目里正常组织均为癌旁正常胃组织，因此将对应的癌旁组织及胃癌组织进行配对比较，分析 COL10A1 在胃癌组织与癌旁组织中的表达差异。

2.1.2. GO、KEGG 富集分析

为探究 COL10A1 在胃癌的发生、发展中所起到的作用，将对分析与 COL10A1 表达差异相关的差异表达基因(Differentially Expression Genes)进行 GO、KEGG 富集分析：根据 COL10A1 的表达量的中位数值作为分界点，按照表达量高低分将胃癌样品分为高表达组及低表达组，然后比较高低表达组之间差异表达的基因，然后将这些基因与 COL10A1 一起进行富集分析。富集分析过程主要是通过使用 cluster Profiler 包进行，通过 GO 分析得到 COL10A1 可能相关的分子功能、生物过程及细胞成分信息，以校正后 p 值 < 0.05 作为过滤条件，选取前十位进行可视化转换；通过 KEGG 分析得到 COL10A1 可能参与的代谢通路，以 p 值小于 0.05 作为过滤条件，选取通路并进行可视化转换。

2.1.3. GSEA 富集分析

GSEA 分析过程同样是建立在根据 COL10A1 表达量高低分组比较得出的差异表达基因的结果之上的。应用 clusterProfiler 包，以由 MSigDB 数据库下载的 C2 基因集作为参考基因集，以 p 值小于 0.05 作为过滤条件，对根据 COL10A1 高低表达分成的高表达组与低表达组数据进行 GSEA 分析。得出在 COL10A1 高表达与低表达时活跃的前 5 条通路，并输出可视化图形结果。

2.1.4. 生存分析

为得到 COL10A1 的表达高低与胃癌患者总生存期(Overall Survival, OS)、无进展生存期(Progression-Free-Survival, PFS)的关系，通过使用 R 软件中的 survival 包对由 TCGA 中下载得到的测序数据以及临床数据进行合并整理，绘制生存曲线图。同时使用 Kaplan-Meierplotter 在线工具分析 COL10A1 表达高低与胃癌患者 OS 与 PFS 之间的关系，并绘制曲线图。

2.1.5. 生存预测模型构建

首先通过使用 survival 包对 COL10A1 进行 COX 单因素及多因素回归分析，明确 COL10A1 表达量的高低对于患者的生存时间是否有影响。如有确认了 COL10A1 表达高低对于患者生存有影响，可通过绘制 ROC 曲线分析 COL10A1 表达对于患者生存预测的能力。ROC 曲线分析是用于评估一个因素预测能力的手段，是可以用于连续型变量分组的方法。在生存分析中，疾病状态和因素取值均会随时间发生变化。而标准的 ROC 曲线分析将个体的疾病状态和因素取值视作固定值，未将时间因素考虑在分析之中。因此使用时间依赖性 ROC 无疑是更好的选择。为进行时间依赖性 ROC 分析，可使用 R 软件中的 timeROC

包分析，分析了在不同时间截点时，COL10A1 对患者的生存预测能力；进一步使用 survival 包、rms 包处理基因表达数据及临床数据构建 COX 风险比例模型，同时将模型图形化，输出列线图(Nomogram)。

2.1.6. 对于肿瘤微环境及免疫细胞浸润影响

为分析 COL10A1 表达对于胃癌肿瘤微环境的影响，首先使用 estimate 包对由 TCGA 下载的肿瘤样本进行分析，得出各个样本的 estimate 评分，然后使用 limma 包将 COL10A1 高表达组与低表达组中的 estimate 评分作比较来评估 COL10A1 对于肿瘤微环境的影响。

为分析 COL10A1 表达对于免疫细胞浸润的影响，首先使用 CIBERSORT 官网提供的 CIBERSORT 包参照同样由官网下载的参考文件对所下载的 TCGA 胃癌样本进行分析处理，得出各个样本中免疫细胞浸润情况，根据 p 值 < 0.05 为条件筛选准确度较高的样本，然后合并基因表达数据与免疫细胞浸润数据，分析 COL10A1 高表达组与低表达组之间各种免疫细胞浸润情况的差异，并进行可视化，输出合并表示的箱型图和浸润与 COL10A1 表达量相关的免疫细胞的相关性图。

2.1.7. 对于免疫治疗效果的影响

为分析 COL10A1 表达量高低对于接受免疫检查点抑制剂治疗的胃癌患者治疗效果的影响，可以通过分析目标基因高低表达组之间免疫治疗效果的差异、目标基因表达量与常见免疫检查点相关基因表达量、目标基因表达量与样本肿瘤突变负荷(Tumor Mutation Burden, TMB)高低等三个方面来分析。

TCIA 数据库是基于 TCGA 数据库中的数据建立的，其中归纳整理了 TCGA 中与免疫相关的数据，其中包括 TCGA-STAD 项目中样本对应的患者接受免疫治疗的效果评价。根据 TCIA 数据库中胃癌患者接受免疫治疗效果的评分数据，分析比较 COL10A1 高表达组与低表达组之间免疫治疗效果评分的差异，从而推测目标基因表达量对于免疫治疗效果的影响。

免疫检查点(Immune Checkpoint)，指程序性死亡受体及其配体。当前胃癌免疫治疗中免疫检查点阻断疗法的基本原理是基于免疫细胞 T 细胞的激活机制。程序性死亡受体表达在 T 细胞表面，其配体表达在肿瘤细胞和髓源性抑制细胞表面。程序性死亡受体与其配体结合可使 T 细胞衰竭而无法正常杀伤肿瘤细胞，肿瘤细胞便可逃脱宿主的免疫监视。因此，基于程序性死亡受体及其配体的免疫检查点阻断疗法通过抑制二者的结合，从而提高宿主免疫系统对肿瘤细胞的攻击性。因此通过分析目标基因与常见免疫检查点相关基因的表达之间的关系，可以推测目标基因与患者预后之间的关系。

肿瘤突变负荷(TMB)被定义为每百万碱基中被检测出的，体细胞基因编码错误、碱基替换、基因插入或缺失错误的总数。肿瘤细胞基因突变频率高，则细胞表面，所携带的肿瘤抗原就多，容易受到机体免疫系统的攻击，而被杀死。因此，检测恶性肿瘤的肿瘤突变负荷，就可以预测患者对免疫治疗的效果，目前已成为临幊上，预测癌症免疫治疗疗效的、具有良好前景的全新生物标志物，在现有的研究中可证实免疫治疗效果与 TMB 水平有关。因此分析 TCGA 下载的胃癌患者的测序数据，得出各样本的 TMB 值，然后比较目标基因高低表达组之间的 TMB 水平差异，分析 COL10A1 表达量与 TMB 值之间的关系。

2.1.8. 对于化疗药物治疗效果的影响

分析常见化疗药物与目标基因表达量之间的关系主要通过 R 软件中的 pRRophetic 包实现。pRRophetic 包是基于其制作者(明尼苏达大学 Paul Gengeher, Nancy Cox, R. Stephanie Huang)所提出的算法：从基因表达数据预测表型，预测外部细胞系(CCLE)的药物敏感性，也可用于临床数据的预测。为分析 COL10A1 表达量高低对于胃癌组织对各种常见抗肿瘤药物敏感性之间的关系，通过使用 R 软件中的 pRRophetic 包处理分析由 TCGA 下载的胃癌患者测序数据，评估 COL10A1 高表达组与低表达组之间药物敏感性差异的分析。

3. 结果

3.1. COL10A1 在多肿瘤及胃癌中表达差异分析

为明确 COL10A1 在多种肿瘤中的表达差异，首先在 UCSC 数据库下载经过 Toil 流程统一处理过的统一标准化的 GTEx 的测序数据，剔除了单个癌种中样本个数小于 3 个的癌种，最终获得了 34 个癌种的表达数据。使用非配对的 Wilcoxon 秩和检验进行差异显著性分析，我们在 25 种肿瘤中观察到了显著上调，在 3 种肿瘤中观察到了显著下调(图 1)。作为验证，使用在线数据库 TIMER2.0 中分析 COL10A1 在多种肿瘤中的表达差异(图 2)。两者结果基本相同。其中胃癌组织与正常胃组织表达具有明显差异。

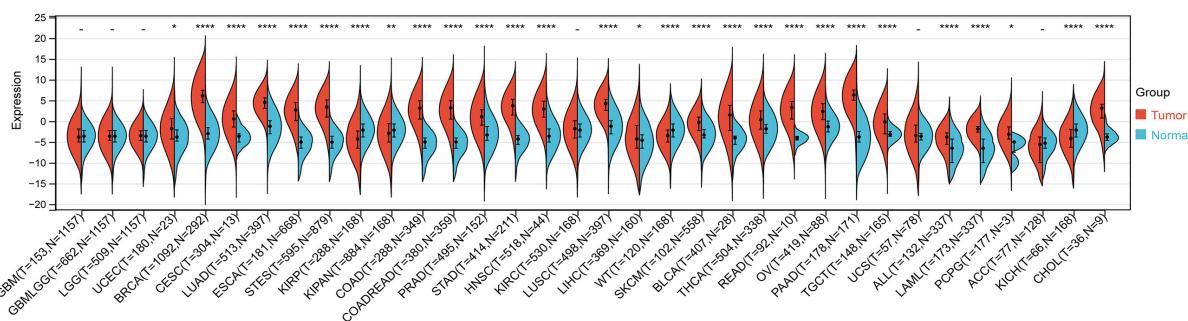


Figure 1. Pancancer analysis of COL10A1 using TCGA combined with GTEx database data

图 1. 使用 TCGA 联合 GTEx 数据库数据对 COL10A1 进行泛癌分析

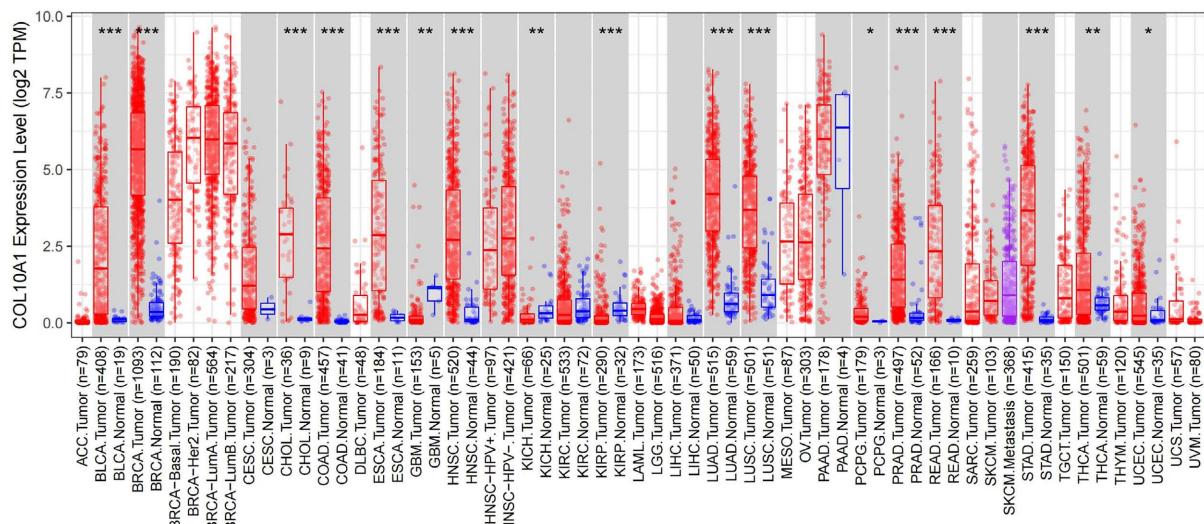


Figure 2. Pan-carcinoma analysis of COL10A1 line based on TCGA data by TIMER2.0 database

图 2. 使用 TIMER2.0 数据库基于 TCGA 数据对 COL10A1 行泛癌分析

为进一步分析 COL10A1 在胃癌组织和正常胃组织之间的表达差异，从 TCGA 数据库的 STAD 项目中下载了共 407 例样本的测序数据，其中共有 375 例样本为胃癌组织测序数据，32 例为癌旁正常胃组织的测序数据。整理提取出各个样本中 COL10A1 的表达量，使用 limma 包对正常组与肿瘤组之间表达量差异(图 3)，将这些正常组织与其对应的肿瘤组织配对比较，并在图中连线表达出来(图 4)。将 GTEx 数据库中的正常胃组织测序数据加入正常组后再行差异分析比较两组中的 COL10A1，结果如图所示(图 5)。综上，正常组与肿瘤组两组中 COL10A1 表达差异明显，在胃癌组织中的表达明显高于正常组织。

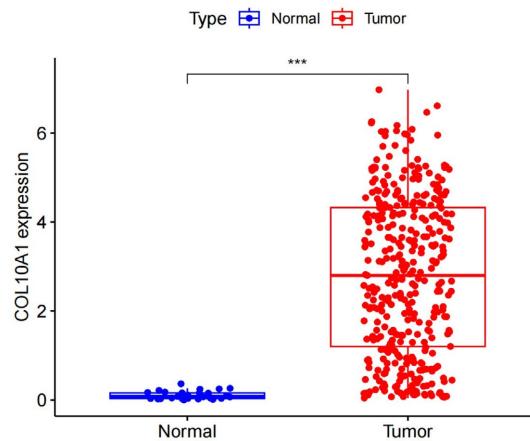


Figure 3. COL10A1 expression comparation in TCGA-STAD data
图 3. 使用 TCGA-STAD 中数据比较 COL10A1 表达量

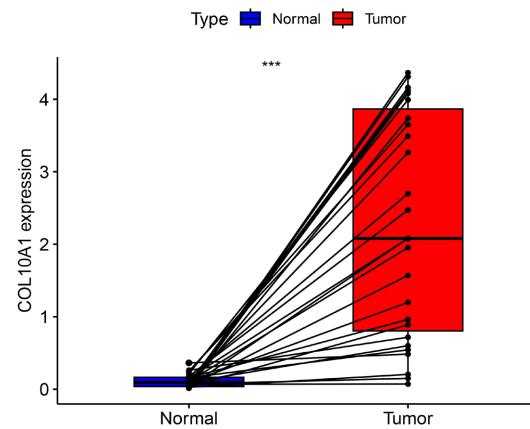


Figure 4. Paired difference analysis on TCGA-STAD data
图 4. 对 TCGA-STAD 数据进行配对差异分析

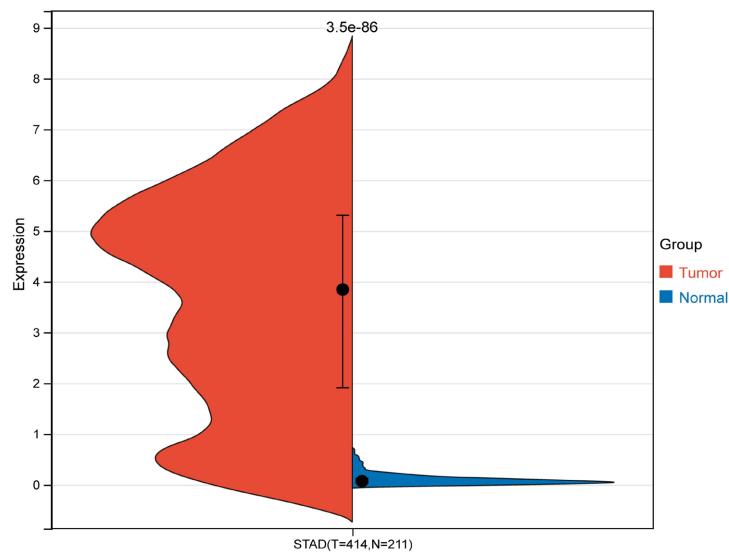


Figure 5. Paired difference analysis on TCGA-STAD data and GTEx data
图 5. 联合 TCGA-STAD 与 GTEx 数据进行差异分析

3.2. GO、KEGG、GSEA 富集分析结果

为将 TCGA 中下载的 STAD 相关测序数据以表达量中位值为筛选条件, 分为高表达组和低表达组, 对高表达组与低表达组之间的表达差异分析, 筛选出可能与 COL10A1 的表达差异相关的基因(DEGs)。将数据分组后对高低表达组之间进行差异分析, 其中以 $|\log FC| > 1$ 且校正后 p 值 < 0.05 作为筛选条件, 共得到差异基因, 其中上调的基因共 242 个、下调的基因共 200 个, 将其中最为差异最为明显的上调基因及下调基因各 50 个的表达差异分析结果可视化, 绘制成为环形图(图 6)及热图(图 7)。

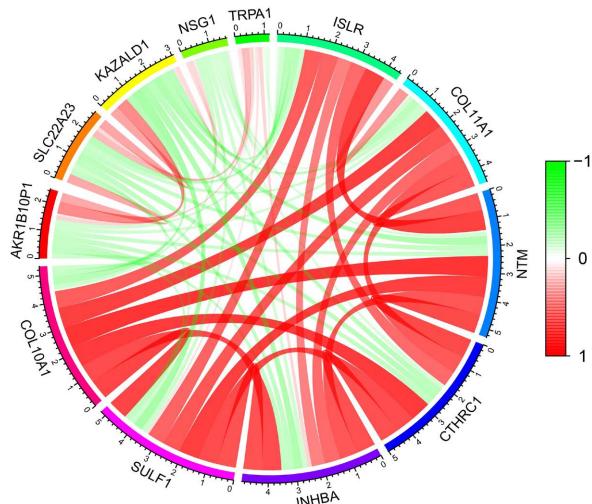


Figure 6. Differential gene analysis results of COL10A1 high and low expression groups (circular graph)

图 6. COL10A1 高低表达组行差异基因分析结果(环形图)

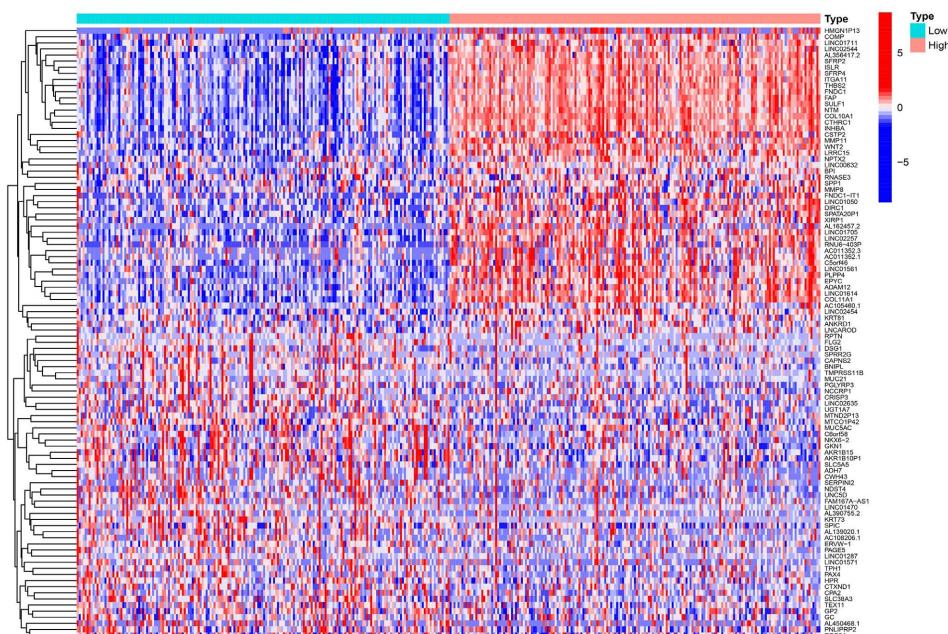


Figure 7. Analysis results of differential genes in COL10A1 high and low expression groups (heat map)

图 7. COL10A1 高低表达组行差异基因分析结果(热图)

将所获得的基因使用 R 软件中的 cluster Profiler 包进行 GO 分析, 其中经 GO 富集分析后共鉴定出可能参与的 284 项基因集, 其中生物过程(biological process, BP)基因集共 213 个, 细胞组分(cellular component, CC)基因集共 21 个, 生物过程(biological process, BP)基因集 50 个, 分别选取三类中可信度最高的 10 中基因集进行可视化转换, 绘制环形图(图 8)及气泡图(图 9)。如图所示, 结果提示主要与细胞外基质、细胞外机构组织、外部封装结构组织等细胞外结构的构成以及骨化、骨骼系统开发、内胚层的功能。

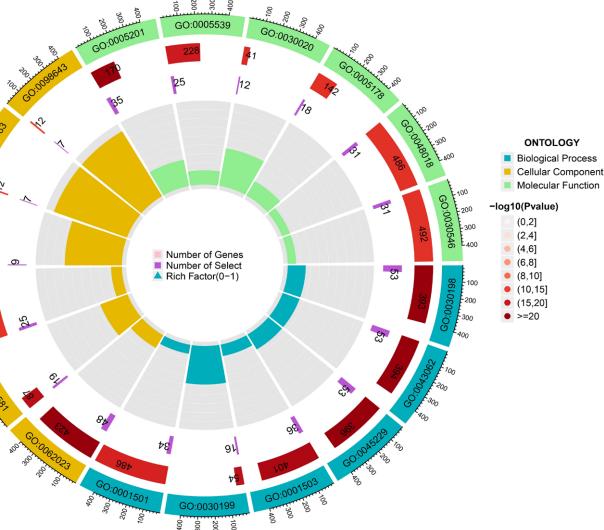


Figure 8. GO enrichment analysis of differentially expressed genes (circular plot)
图 8. 对差异表达基因行 GO 富集分析(环形图)

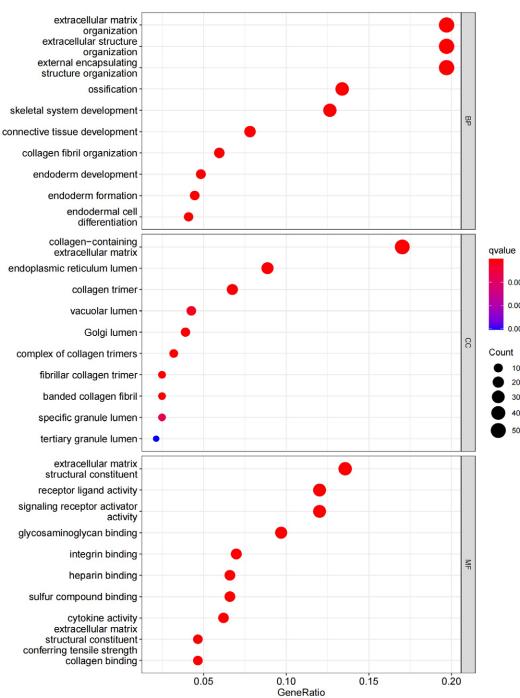


Figure 9. GO enrichment analysis of differentially expressed genes (bubble plot)
图 9. 对差异表达基因行 GO 富集分析(气泡图)

使用 clusterProfiler 包对 DEGs 进行 KEGG 富集分析，依据筛选条件共选出可信度最高的 18 个通路信息，将富集分析结果进行可视化转换，绘制条形图(图 10)。根据结果所示可信度排名较前的几条通路($p < 0.01$)主要与以下几个功能有关：1、细胞间信息传递；2、IL-17 信号通路；3、TGF- β 信号通路；4、PI3K-Akt 信号通路。

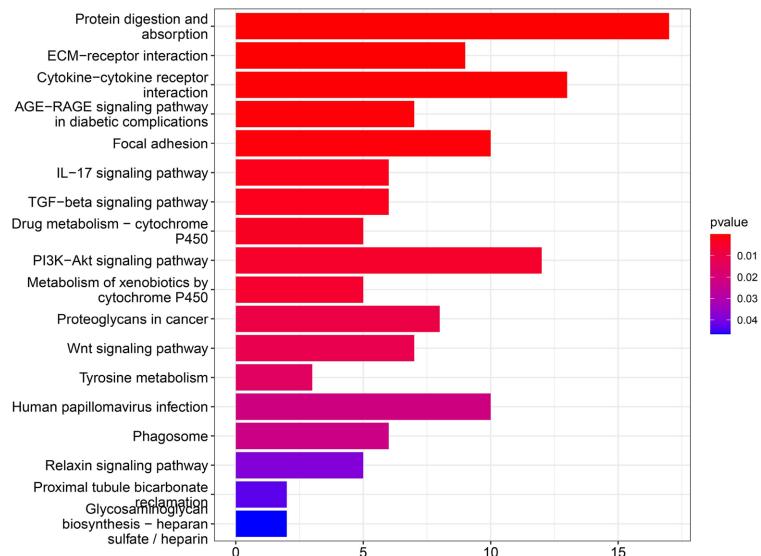


Figure 10. KEGG analysis results of differentially expressed genes
图 10. 对差异表达基因进行 KEGG 分析结果

使用 clusterProfiler 包，以 Msigdb 数据库下载的 C2 集中的 c2.cp.kegg.v7.4.symbols 基因集作为参考基因集，对筛选出的 DEGs 进行 GSEA 分析，共得到可能参与的通路共 37 条，其中满足筛选条件的通路共 14 条，选择其中可信度最高的 5 条通路进行可视化转换，结果如图(图 11)。如图所示随目标基因高表达激活的通路是与细胞间信息传递功能相关通路与 TGF- β 信号通路，受到抑制的为嗅觉传导相关通路。

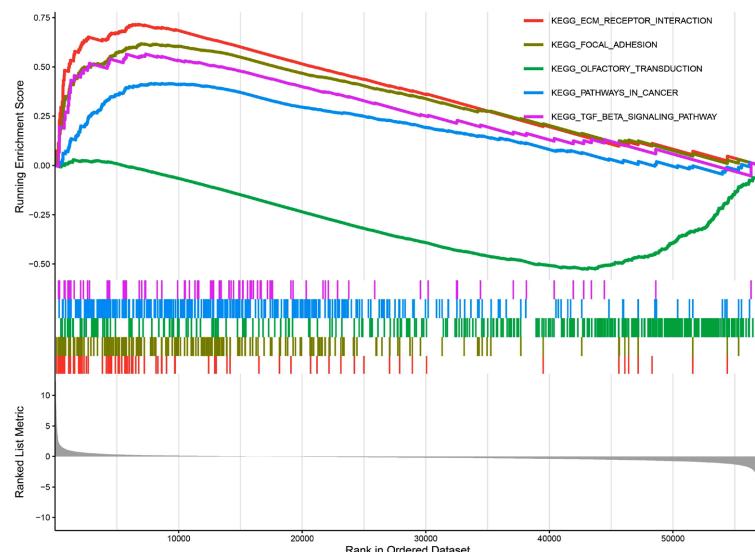


Figure 11. GSEA analysis results of differentially expressed genes
图 11. 对差异表达基因行 GSEA 分析结果

3.3. 生存分析结果

将胃癌样本的测序数据与临床数据整合, 去除掉测序数据中正常组织部分, 然后根据 COL10A1 的表达量中位值, 将样本分为高表达组及低表达组。然后使用 limma、survival 和 survminer 包来分析整合数据: 比较高、低表达组之间的生存差异得到显著性 p 值并绘制总生存期(Overall Survival, OS)曲线(图 12)。以同样流程得出无进展生存期(Progression-Free-Survival, PFS) (图 13)。同时作为验证, 使用 Kaplan-Meierplotter 数据库中的在线分析工具, 绘制 OS 曲线(图 14)及 PFS 曲线(由于 K-M plotter 数据库没有 PFS 选项, 因此使用 FirstProgression 曲线代替)(图 15)。如图 TCGA 数据库 STAD 数据所得出的 OS 曲线图和 K-Mplotter 数据库所得到的 OS 曲线图所示, 高、低表达组之间生存曲线之间 p 值均小于 0.05, 而 PFS 图中 p 值均大于 0.05。说明 COL10A1 表达高低对于胃癌患者的总生存时间有影响, 高表达患者总生存时间较短, 而 COL10A1 表达量高低表达组之间无进展生存期长短无差异, 对于患者的无进展生存期无影响。

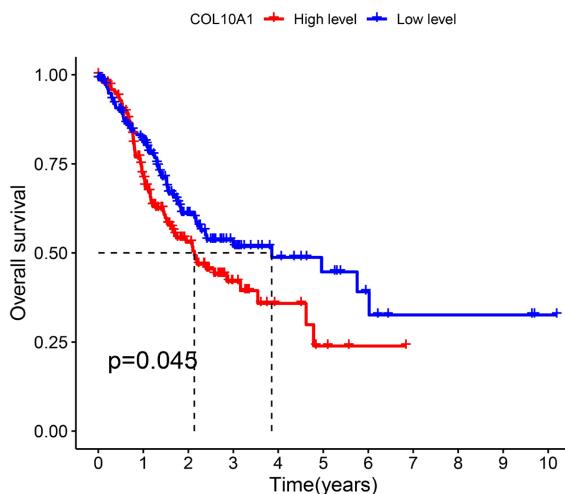


Figure 12. Drawing OS curve based on TCGA-STAD data
图 12. 基于 TCGA-STAD 数据绘制 OS 曲线

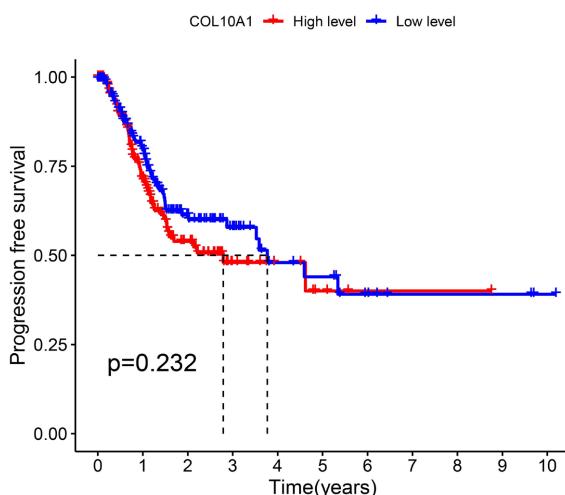


Figure 13. Drawing PBS curve based on TCGA-STAD data
图 13. 基于 TCGA-STAD 数据绘制 PBS 曲线

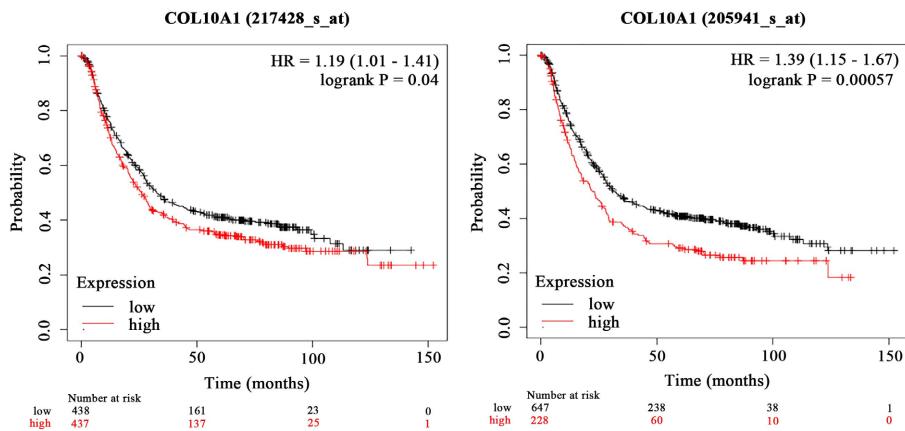


Figure 14. Drawing OS Curve by using Kaplan Meier Plotter
图 14. 使用 Kaplan-Meier plotter 绘制 OS 曲线

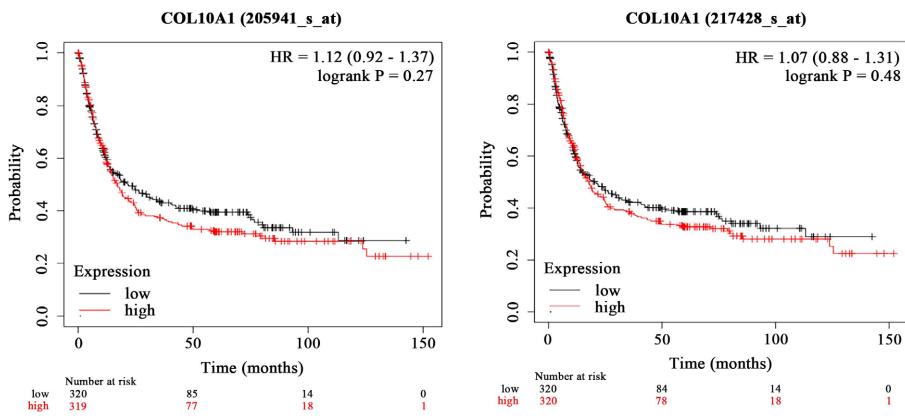


Figure 15. Drawing FP Curve by using Kaplan Meier Plotter
图 15. 使用 Kaplan-Meier plotter 绘制 FP 曲线

3.4. 生存预测模型构建

为探索 COL10A1 表达量与患者的生存状态之间的关系，首先处理由 TCGA 下载的临床数据，将数据根据目标基因的表达量高低将样本分为两组，分别分析两组之间各临床数据之间的差别，结果如图所示(图 16、图 17)，两组样本之间年龄、性别、肿瘤 Grade 分期、M 和 N 分期等信息无明显差别，Stage 分期有明显差异，T 分期有显著差异。

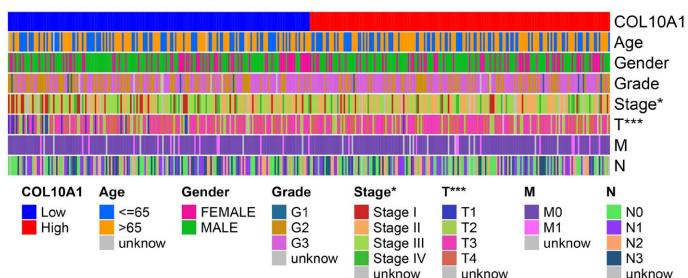


Figure 16. Differences in clinical characteristics between high and low expression groups of COL10A1
图 16. COL10A1 高低表达组之间临床特征差异

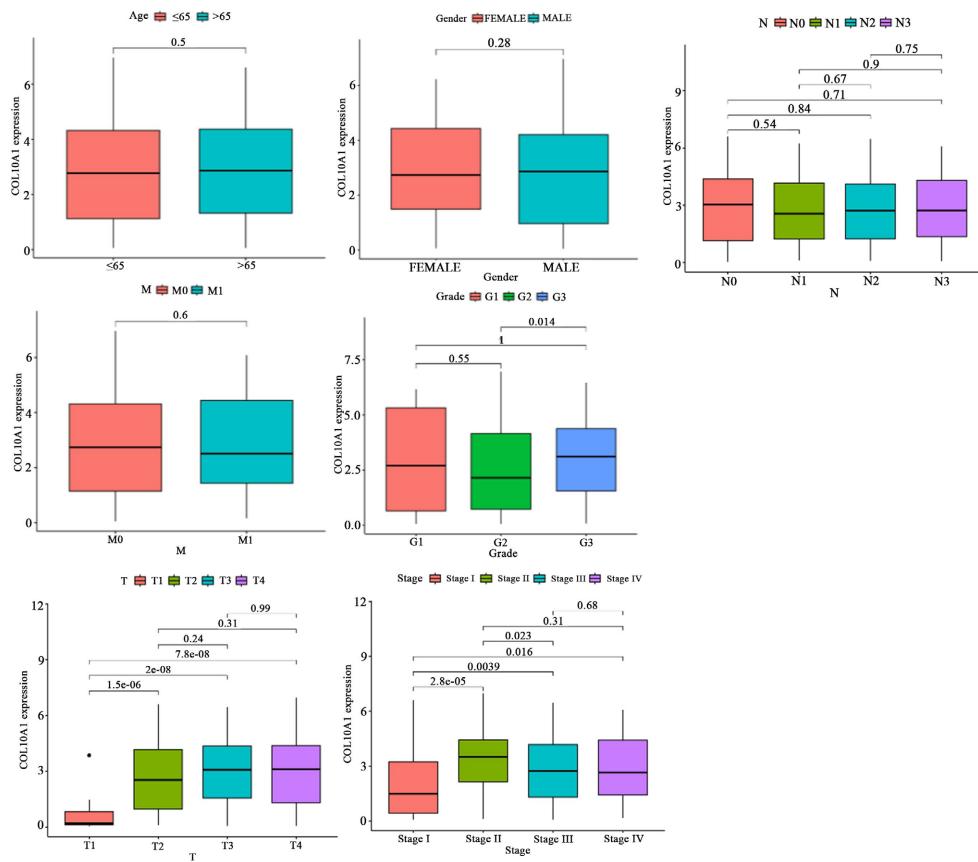


Figure 17. Differences in clinical features between high and low expression groups
图 17. 高低表达组之间各个临床特征差异

使用 survival 包对临床数据进行分析, 从而明确 COL10A1 的表达量以及年龄、TNM 分期、Grade 分级和 Stage 分期等其他临床数据对生存的影响。首先对上述的各种因素行单因素 COX 分析, 将结果绘制成森林图表示(图 18), 如图所示 COL10A1 表达量高低在单因素 COX 分析中 p 值小于 0.05, HR 大于 1; 同样步骤将上述各影响因素行多因素 COX 分析(图 19), 如图所示 COL10A1 表达量高低在分析中的 p 值仍小于 0.05 且 HR 大于 1。综上可初步判定 COL10A1 的表达量高低与患者的预后有关。

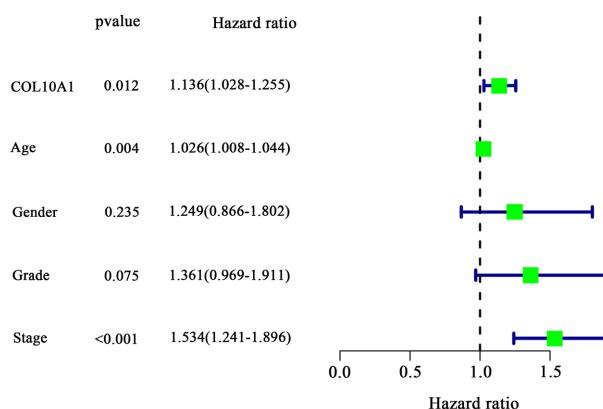
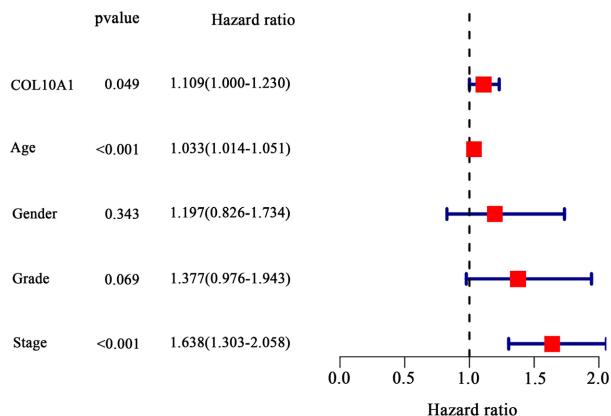
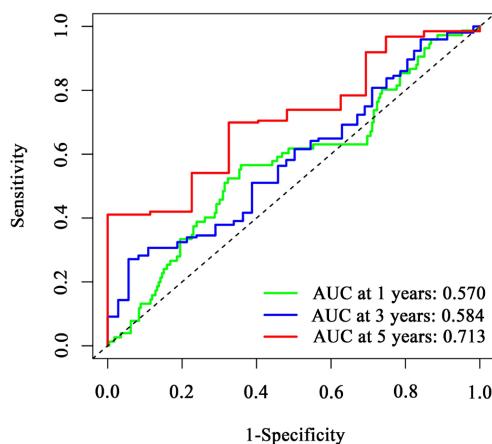


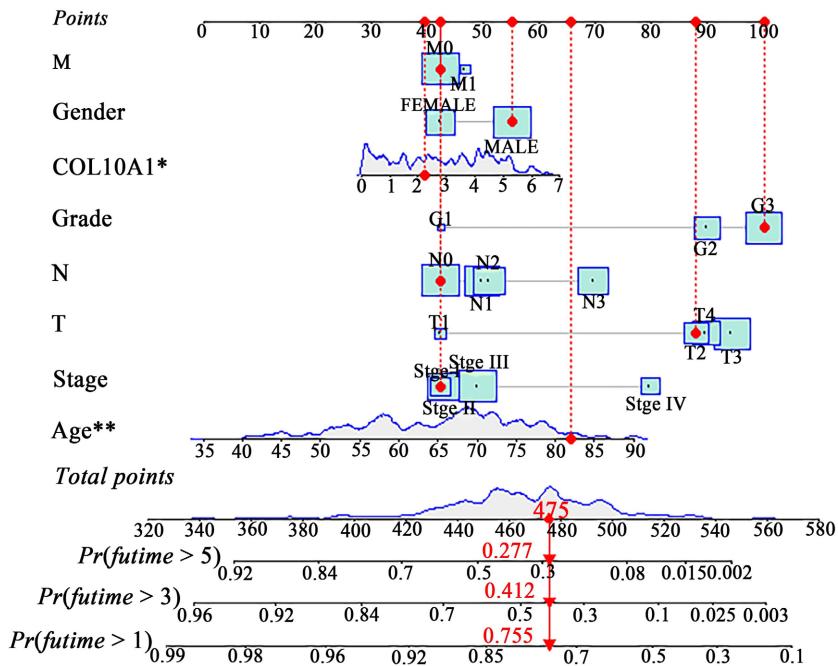
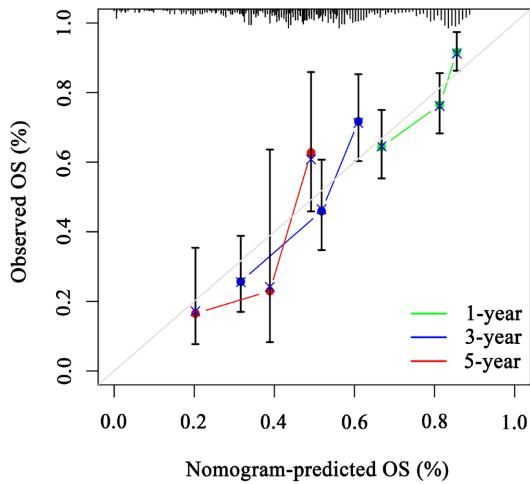
Figure 18. Single factor Cox analysis results
图 18. 单因素 cox 分析结果

**Figure 19.** Multifactor cox analysis results**图 19.** 多因素 cox 分析结果

使用 timeROC 包分析临床数据绘制时间依赖性 ROC 曲线，评估 1 年、3 年、5 年时 COL10A1 表达量对于患者生存预测能力，曲线如图所示(图 20)，5 年时曲线下面积最大，为 0.713，而 1 年、3 年时曲线下面积分别为 0.570、0.584，提示 COL10A1 在 1 年、3 年时具有一定的生存预测能力，而 5 年时生存预测能力最佳。

**Figure 20.** Time-ROC curve**图 20.** 时间依赖性 ROC 曲线

进一步使用 R 软件中 survival、rms 包使用 COL10A1、年龄、性别、肿瘤分期等因素行多因素 COX 回归分析，构建 COX 比例风险模型，使用 R 软件的 regplot 包将该模型以列线图(nomogram) (图 21) 的形式表现出来，同时输出预测模型的校正曲线。该图的内容主要有：1) 预测模型中的变量(每一个变量上都有其可取值范围)；2) 得分(每个影响因素在不同的取值下都具有其对应的单项得分，以及所有变量对应的总得分)；3) 预测概率(总得分对应到 1 年、3 年、5 年上的生存概率)。如图所示，假如患者总分为 475 分，则该患者 1 年生存率为 75.5%、3 年生存率为 41.2%、5 年生存率为 27.7%。该模型共纳入了年龄、性别、TNM 分期、Grade 分期、Stage 分期等因素，其中年龄的影响度较高，其次为 COL10A1 表达量。与列线图同时得出的校准曲线(calibrationcurve) (图 22)，如图所示三条线分别为 1 年、3 年、5 年预测值的曲线，基本沿对角线分布，预测效果尚可。

**Figure 21.** Multivariate cox regression prognostic model**图 21.** 多变量 cox 回归预后预测模型**Figure 22.** Prediction model correction curve**图 22.** 预测模型矫正曲线

3.5. COL10A1 对于胃癌肿瘤微环境的影响

使用 estimate 包对由 TCGA 下载的肿瘤样本进行分析, 得出每个样本的基质得分(Stromal Score)和免疫得分(Immune Score), 由两者联合得出该样本的 Estimate 评分(ESTIMATE Score), 并得出每个样本的肿瘤纯度。比较高、低表达组之间基质得分、免疫得分和 Estimate 评分的差异, 如图所示, 将差异分析结果绘制成小提琴图(图 23), 如图所示, 基质得分和 Estimate 评分在高、低表达组之间具有显著的差异, p 值小于 0.001; 免疫评分在高、低表达组之间也有明显差异, 其 p 值小于 0.05。综上, COL10A1 的高表达者胃癌组织的 Estimate 评分较高, 也意味着该患者胃癌组织纯度更低。

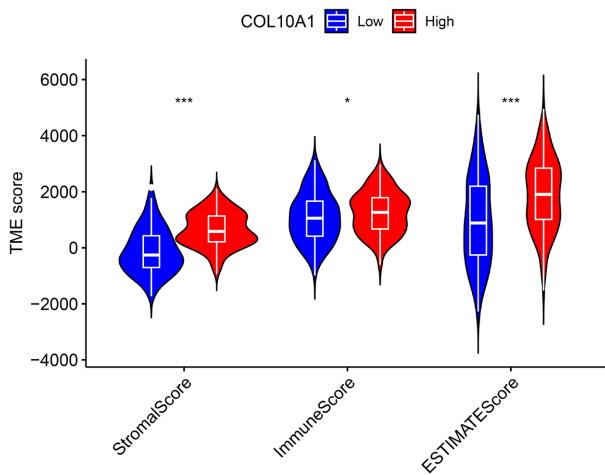


Figure 23. Comparison of matrix score, immune score, and ESTIMATE score between high and low expression groups

图 23. 高低表达组之间基质评分、免疫评分及 ESTIMATE 评分比较

3.6. COL10A1 对于胃癌组织免疫细胞浸润影响

使用 R 软件中由 Cibersort 官方提供的 Cibersort 包对由 TCGA 下载的胃癌测序数据进行整理、分析。根据目标基因 COL10A1 的表达量高低将该样本分为高、低表达组，然后使用 limma 包比较高、低表达组之间免疫细胞的差异，并将结果转化为箱型图(图 24)。对目标基因表达量以及免疫细胞表达量之前做相关性分析，结果以 Lolipop 图(图 25)表达出来，如图所示，如图所示，M2 型巨噬细胞、B 记忆细胞在两组中的表达差异显著，其 p 值小于 0.001；激活的树突状细胞在两组中表达差异较明显，p 值小于 0.01；M0、M1 型巨噬细胞，调节性 T 细胞在两组中有一定差异。可推断 COL10A1 的高表达可能与免疫细胞大量浸润有关。

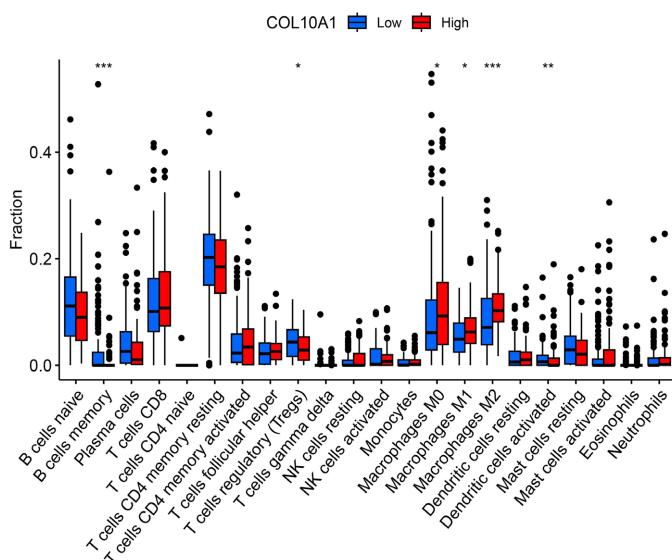


Figure 24. Comparison of immune cell infiltration between high and low expression groups (box plot)

图 24. 高低表达组之间免疫细胞浸润程度比较(箱型图)

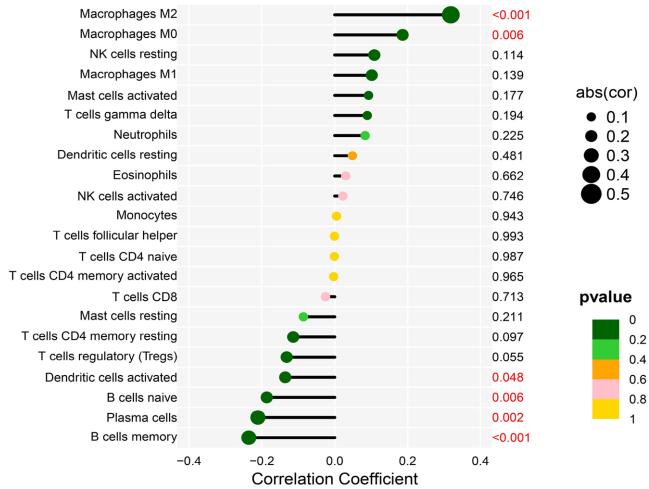


Figure 25. Comparison of immune cell infiltration between high and low expression groups (lollipop diagram)

图 25. 高低表达组之间免疫细胞浸润程度比较(棒棒糖图)

3.7. COL10A1 对于胃癌患者免疫治疗效果的影响

将TCIA数据库下载的免疫治疗打分文件与下载的TCGA中胃癌测序数据整合,根据样本ID将TCIA中关于免疫治疗效果打分文件与整理后的测序数据整合起来。根据COL10A1的表达量分为高、低表达两组,使用limma包将高、低表达组之间免疫治疗评分进行比较,评估高低表达组在CTLA4阳性、PD-1阴性,CTLA4阴性、PD-1阳性,CTLA4阴性、PD-1阴性,CTLA4阳性、PD-1阳性等四种情况下免疫治疗效果,并使用ggpurr包将最终差异比较结果转换为小提琴图表达出来(图26),如图所示,在CTLA4阳性、PD-1阴性,以及CTLA4阴性、PD-1阴性这两组中,高、低表达组之间的免疫治疗评分具有显著差距。

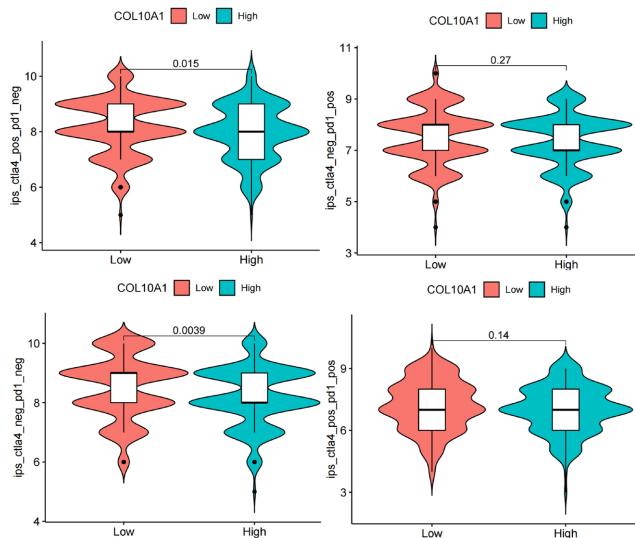


Figure 26. Differences in immunotherapy efficacy between high and low expression groups after grouping based on the expression of CTLA4 and PD-1

图 26. 根据 CTLA4 与 PD-1 表达情况分组后, 高低表达组之间免疫治疗效果差异

为探究目标基因表达量与常见免疫检查点相关基因之间的关系，提取肿瘤样本测序数据中的免疫检查点相关基因的表达量，使用皮尔森相关系数对这些免疫检查点基因表达量与目标基因表达量进行相关性检验，相关性检验过滤条件设置为 p 值 < 0.001 ，将分析结果转化为热力图(图 27)，如图所示，目标基因表达与 HAVCR2、NRP1、CD70、TNFSF4、PDCD1LG2、CD86、LAIR1、TNFSF9、CD276、TNFSF18、TNFRSF4、CD200 等多种免疫检查点之间为正调控关系。

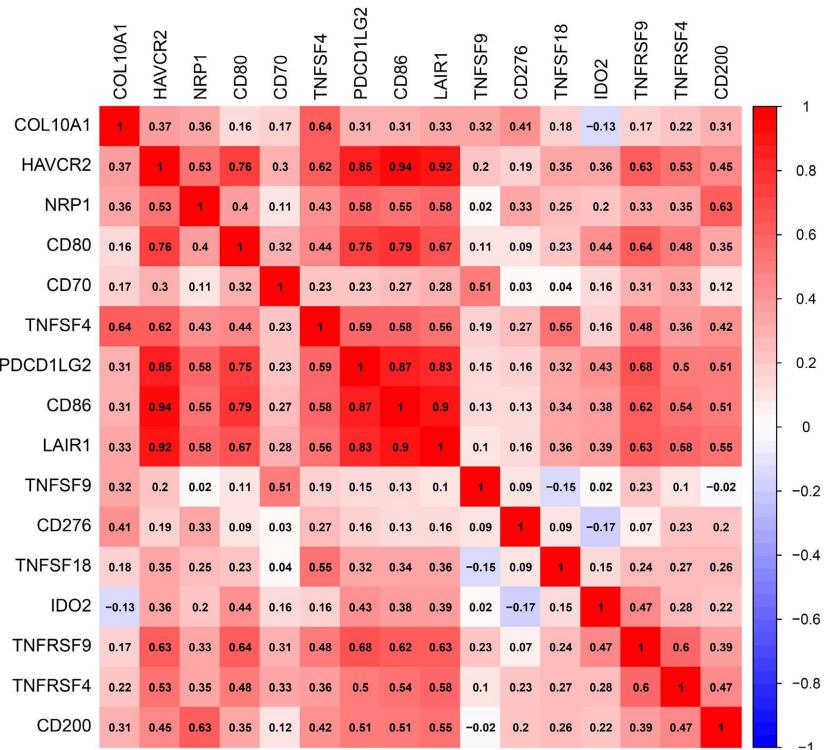


Figure 27. Relationship between the expression level of target gene COL10A1 and common immune checkpoint expression levels

图 27. 目标基因 COL10A1 表达量与常见免疫检查点表达量关系

4. 讨论

在本研究中，为了确定 COL10A1 在胃癌中的作用，首先是通过 TIMER2.0 及联合 TCGA 联合 GTEx 数据在多种瘤中进行差异表达分析，根据分析结果在除了在膀胱癌、乳腺癌、头颈部鳞癌、肾乳头状细胞癌、肺鳞癌、肺腺癌、前列腺癌等肿瘤中肿瘤组织与正常组织中的目标基因表达量有显著差异，同时在胃癌、直肠癌、胆管癌、结肠癌、食管癌等消化道相关的肿瘤中，肿瘤与正常组织中具有显著的表达差异，其中结直肠癌[14]、食管癌、乳腺癌[15]、肺癌[17] [26]中的差异已有研究证明。通过绘制根据目标基因表达量高低的分组中的胃癌患者的总生存期和无进展生存期曲线的差异，可以得知目标基因高表达组的总生存期相较于低表达组的生存期更低，而在无进展生存期中高低表达组之间无明显差异。进一步通过单因素、多因素 COX 回归分析发现 COL10A1 的高表达、年龄、高 Stage 分期等是胃癌患者的预后独立因素。通过汇总上述预后相关因素，构建了对于患者预后的预测模型，通过绘制时间依赖性的 ROC 曲线及矫正曲线进行评估，该模型对于胃癌患者生存具有一定的预测效果。同时根据已有研究显示，胃癌患者血液中的 COL10A1 表达量亦有显著差距，且与胃癌患者的预后有关，具有成为胃癌预后的标志物潜力[27]，该结果与本研究结果可以互相支持验证。

对于免疫治疗，当前已证实的是免疫微环境特征可以作为评估肿瘤细胞对于免疫治疗的反应及影响的可靠的生物标志物之一。Estimate 算法通过对基因表达数据分析，来预测恶性肿瘤组织中的间质细胞和免疫细胞的含量，推测肿瘤纯度，从而评估肿瘤微环境情况。本研究中通过使用 R 软件中的 estimate 包为胃癌样本进行打分，然后比较目标基因高低表达组之间的差异，结果显示高表达组中的样本三项得分均显著高于低表达组，预示着目标基因的表达与肿瘤微环境中的基质和免疫细胞呈正相关。进一步使用 Cibersort 算法进一步探究 COL10A1 表达与各种免疫细胞的浸润水平之间的关系，结果提示 COL10A1 高表达组具有更多的免疫细胞浸润，其中 M0、M1、M2 巨噬细胞，未活化的树突状细胞的浸润水平成正相关。此外，结果表明 COL10A1 与 HAVCR2、NRP1、CD70、TNFSF、PDCD1LG2、CD86、LAIR1、TNFSF9、CD276、TNFSF18、TNFRSF4 及 CD200 常见的免疫检查点基因的表达量成正相关。同时通过分析目标基因 COL10A1 高低表达组之间患者对于免疫治疗的效果评分，表明高表达组的患者应用免疫治疗效果较差。目前根据现有研究表明，细胞外基质与功能失调的免疫逃逸与免疫治疗的失败密切相关。综上，我们猜测 COL10A1 的高表达可能通过影响肿瘤微环境的构成及免疫检查点的表达导致胃癌患者接受免疫治疗效果较差，提示 COL10A1 具有成为潜在的免疫治疗靶点以及成为免疫检查点阻断治疗的预测指标。

通过对目标基因高低表达组之间的差异表达基因的 GO、KEGG 和 GSEA 分析提示 COL10A1 可能参与了 TGF- β 信号通路[28]、PI3K-Akt 信号通路[29]。目前两种信号通路已经经过长时间研究：TGF- β 信号通路在成熟有机体和发育中的胚胎中都参与了许多细胞过程，这些过程包括细胞生长，细胞分化，细胞凋亡，细胞动态平衡等其它细胞功能，目前大量研究表明该通路在肿瘤的进展和转移中起到关键作用；PI3K-AKT 是细胞中一个经典的信号通路，它是一条与磷脂酰肌醇有关的信号通路，也是 RTK 介导衍生的信号通路。在胰岛素的刺激以及细胞生存当中发挥着重要作用；AKT 作为一种原癌基因，在多种细胞过程中发挥关键作用，如葡萄糖代谢，细胞凋亡，细胞增殖，转录和细胞迁移；PI3K/AKT 信号通路失调见于多种人类疾病，包括癌症、糖尿病、心血管疾病和神经疾病。这些潜在的信号通路为后续继续对目标基因 COL10A1 的基质研究提供了参考。

综上所述，在本研究中通过分析目标基因 COL10A1 在胃癌细胞中的高表达，并与胃癌患者的预后较差存在显著相关，通过该基因以及年龄、Stage 分期等其他因素所构成的临床预测模型可较好的预测患者 1 年、3 年、5 年的总生存期。对于目标基因可能参与的机制的探索，该基因在胃癌中与免疫细胞的浸润和免疫检查点的表达具有显著性相关，并且可能通过 TGF- β 信号通路、PI3K-Akt 信号通路等促进胃癌的发生与发展。结合分析表明目标基因 COL10A1 具有作为胃癌患者免疫治疗靶点的潜在能力。

5. 结论

COL10A1 在胃癌中高表达，当 COL10A1 表达上调时胃癌患者的总体生存期显著缩短，但对无进展生存期无明显影响，提示其与患者的预后不良有关。COL10A1 与肿瘤微环境的构成有关，与免疫细胞浸润和免疫检查点的表达存在显著性相关。通过分析 TCIA 数据库中数据，COL10A1 高表达会使部分患者的免疫治疗效果差，提示 COL10A1 可能是一个潜在的免疫治疗靶点。

参考文献

- [1] Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021) Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, **71**, 209-249. <https://doi.org/10.3322/caac.21660>
- [2] Xia, C., Dong, X., Li, H., Cao, M., Sun, D., He, S., et al. (2022) Cancer Statistics in China and United States, 2022: Profiles, Trends, and Determinants. *Chinese Medical Journal*, **135**, 584-590.

- <https://doi.org/10.1097/CM9.0000000000002108>
- [3] Smyth, E.C., Nilsson, M., Grabsch, H.I., van Grieken, N.C. and Lordick, F. (2020) Gastric Cancer. *The Lancet*, **396**, 635-648. [https://doi.org/10.1016/S0140-6736\(20\)31288-5](https://doi.org/10.1016/S0140-6736(20)31288-5)
- [4] Muro, K., Van Cutsem, E., Narita, Y., Pentheroudakis, G., Baba, E., Li, J., et al. (2019) Pan-Asian Adapted ESMO Clinical Practice Guidelines for the Management of Patients with Metastatic Gastric Cancer: A JSMO-ESMO Initiative Endorsed by CSCO, KSMO, MOS, SSO and TOS. *Annals of Oncology*, **30**, 19-33. <https://doi.org/10.1093/annonc/mdy502>
- [5] Smyth, E.C. and Cervantes, A. (2020) Addition of Nivolumab to Chemotherapy in Patients with Advanced Gastric Cancer: A Relevant Step Ahead, But Still Many Questions to Answer. *ESMO Open*, **5**, E001107. <https://doi.org/10.1136/esmopen-2020-001107>
- [6] Rizzo, A., Mollica, V., Ricci, A.D., Maggio, I., Massucci, M., Rojas Limpe, F.L., et al. (2020) Third- and Later-Line Treatment in Advanced or Metastatic Gastric Cancer: A Systematic Review and Meta-Analysis. *Future Oncology*, **16**, 4409-4418. <https://doi.org/10.2217/fon-2019-0429>
- [7] Bang, Y.J., Kang, Y.K., Catenacci, D.V., Muro, K., Fuchs, C.S., Geva, R., et al. (2019) Pembrolizumab Alone or in Combination with Chemotherapy as First-Line Therapy for Patients with Advanced Gastric or Gastroesophageal Junction Adenocarcinoma: Results from the Phase II Nonrandomized KEYNOTE-059 Study. *Gastric Cancer*, **22**, 828-837. <https://doi.org/10.1007/s10120-018-00909-5>
- [8] Liu, T., Bai, Y., Lin, X., Li, W., Wang, J., Zhang, X., et al. (2023) First-Line Nivolumab plus Chemotherapy vs Chemotherapy in Patients with Advanced Gastric, Gastroesophageal Junction and Esophageal Adenocarcinoma: Check-Mate 649 Chinese Subgroup Analysis. *International Journal of Cancer*, **152**, 749-760. <https://doi.org/10.1002/ijc.34296>
- [9] He, Y., Siebuhr, A.S., Brandt-Hansen, N.U., Wang, J., Su, D., Zheng, Q., et al. (2014) Type X Collagen Levels Are Elevated in Serum from Human Osteoarthritis Patients and Associated with Biomarkers of Cartilage Degradation and Inflammation. *BMC Musculoskeletal Disorders*, **15**, Article No. 309. <https://doi.org/10.1186/1471-2474-15-309>
- [10] Gu, J., Lu, Y., Li, F., Qiao, L., Wang, Q., Li, N., et al. (2014) Identification and Characterization of the Novel Col10a1 Regulatory Mechanism during Chondrocyte Hypertrophic Differentiation. *Cell Death & Disease*, **5**, e1469. <https://doi.org/10.1038/cddis.2014.444>
- [11] Hasegawa, K., Higuchi, Y., Yamashita, M. and Tanaka, H. (2015) Japanese Familial Case with Metaphyseal Dysplasia, Schmid Type Caused by the p.T555P Mutation in the COL10A1 Gene. *Clinical Pediatric Endocrinology*, **24**, 33-36. <https://doi.org/10.1297/cpe.24.33>
- [12] Boudreau, N. and Bissell, M.J. (1998) Extracellular Matrix Signaling: Integration of Form and Function in Normal and Malignant Cells. *Current Opinion in Cell Biology*, **10**, 640-646. [https://doi.org/10.1016/S0955-0674\(98\)80040-9](https://doi.org/10.1016/S0955-0674(98)80040-9)
- [13] Sole, X., Crous-Bou, M., Cordero, D., Olivares, D., Guino, E., Sanz-Pamplona, R., et al. (2014) Discovery and Validation of New Potential Biomarkers for Early Detection of Colon Cancer. *PLOS ONE*, **9**, e106748. <https://doi.org/10.1371/journal.pone.0106748>
- [14] Kahlert, U.D., Shi, W., Strecker, M., Scherpinski, L.A., Wartmann, T., Dolling, M., et al. (2022) COL10A1 Allows Stratification of Invasiveness of Colon Cancer and Associates to Extracellular Matrix and Immune Cell Enrichment in the Tumor Parenchyma. *Frontiers in Oncology*, **12**, Article1007514. <https://doi.org/10.3389/fonc.2022.1007514>
- [15] Giussani, M., Landoni, E., Merlini, G., Turdo, F., Veneroni, S., Paolini, B., et al. (2018) Extracellular Matrix Proteins as Diagnostic Markers of Breast Carcinoma. *Journal of Cellular Physiology*, **233**, 6280-6290. <https://doi.org/10.1002/jcp.26513>
- [16] Zhang, M., Chen, H., Wang, M., Bai, F. and Wu, K. (2020) Bioinformatics Analysis of Prognostic Significance of COL10A1 in Breast Cancer. *Bioscience Reports*, **40**, BSR20193286. <https://doi.org/10.1042/BSR20193286>
- [17] Guo, Q., Zheng, M., Xu, Y., Wang, N. and Zhao, W. (2019) MiR-384 Induces Apoptosis and Autophagy of Non-Small Cell Lung Cancer Cells through the Negative Regulation of Collagen α -1(X) Chain Gene. *Bioscience Reports*, **39**, BSR20181523. <https://doi.org/10.1042/BSR20181523>
- [18] Hinshaw, D.C. and Shevde, L.A. (2019) The Tumor Microenvironment Innately Modulates Cancer Progression. *Cancer Research*, **79**, 4557-4566. <https://doi.org/10.1158/0008-5472.CAN-18-3962>
- [19] Bense, R.D., Sotiriou, C., Piccart-Gebhart, M.J., Haanen, J., van Vugt, M., de Vries, E.G.E., et al. (2017) Relevance of Tumor-Infiltrating Immune Cell Composition and Functionality for Disease Outcome in Breast Cancer. *Journal of the National Cancer Institute*, **109**, djw192. <https://doi.org/10.1093/jnci/djw192>
- [20] Yang, Y., He, X., Tang, Q.Q., Shao, Y.C., Song, W.J., Gong, P.J., et al. (2021) GMFG Has Potential to Be a Novel Prognostic Marker and Related to Immune Infiltrates in Breast Cancer. *Frontiers in Oncology*, **11**, Article 629633. <https://doi.org/10.3389/fonc.2021.629633>
- [21] Li, J., Chen, D. and Shen, M. (2022) Tumor Microenvironment Shapes Colorectal Cancer Progression, Metastasis, and

- Treatment Responses. *Frontiers in Medicine*, **9**, Article 869010. <https://doi.org/10.3389/fmed.2022.869010>
- [22] Chen, K., Xing, J., Yu, W., Xia, Y., Zhang, Y., Cheng, F., et al. (2021) Identification and Validation of Hub Genes Associated with Bladder Cancer by Integrated Bioinformatics and Experimental Assays. *Frontiers in Oncology*, **11**, Article 782981. <https://doi.org/10.3389/fonc.2021.782981>
- [23] Du, Y., Jiang, X., Wang, B., Cao, J., Wang, Y., Yu, J., et al. (2021) The Cancer-Associated Fibroblasts Related Gene CALD1 Is a Prognostic Biomarker and Correlated with Immune Infiltration in Bladder Cancer. *Cancer Cell International*, **21**, Article No. 283. <https://doi.org/10.1186/s12935-021-01896-x>
- [24] Huang, H., Zhu, L., Huang, C., Dong, Y., Fan, L., Tao, L., et al. (2021) Identification of Hub Genes Associated with Clear Cell Renal Cell Carcinoma by Integrated Bioinformatics Analysis. *Frontiers in Oncology*, **11**, Article 726655. <https://doi.org/10.3389/fonc.2021.726655>
- [25] Yoshihara, K., Shahmoradgoli, M., Martinez, E., Vegesna, R., Kim, H., Torres-Garcia, W., et al. (2013) Inferring Tumour Purity and Stromal and Immune Cell Admixture from Expression Data. *Nature Communications*, **4**, Article No. 2612. <https://doi.org/10.1038/ncomms3612>
- [26] Andriani, F., Landoni, E., Mensah, M., Facchinetto, F., Miceli, R., Tagliabue, E., et al. (2018) Diagnostic Role of Circulating Extracellular Matrix-Related Proteins in Non-Small Cell Lung Cancer. *BMC Cancer*, **18**, Article No. 899. <https://doi.org/10.1186/s12885-018-4772-0>
- [27] Necula, L., Matei, L., Dragu, D., Pitica, I., Neagu, A.I., Bleotu, C., et al. (2020) High Plasma Levels of COL10A1 Are Associated with Advanced Tumor Stage in Gastric Cancer Patients. *World Journal of Gastroenterology*, **26**, 3024-3033. <https://doi.org/10.3748/wjg.v26.i22.3024>
- [28] Caja, L., Dituri, F., Mancarella, S., Caballero-Diaz, D., Moustakas, A., Giannelli, G., et al. (2018) TGF- β and the Tissue Microenvironment: Relevance in Fibrosis and Cancer. *International Journal of Molecular Science*, **19**, Article 1294. <https://doi.org/10.3390/ijms19051294>
- [29] Noorolyai, S., Shajari, N., Baghbani, E., Sadreddini, S. and Baradaran, B. (2019) The Relation between PI3K/AKT Signalling Pathway and Cancer. *Gene*, **698**, 120-128. <https://doi.org/10.1016/j.gene.2019.02.076>