· 数据挖掘与循证医学 ·

应用机器学习和神经网络模型识别结直肠癌“炎癌转化”过程的关键基因及防治中药预测

刘　洋[1]，曹赛雅[1]，冯月娇[2]，王　杰[1,2]，陈　腾[1*]

1. 上海中医药大学附属普陀医院（上海市普陀区中心医院） 普外科，上海　200062
2. 安徽医科大学普陀中心临床学院 普外科，上海　200062

摘　要：目的　应用公共数据库挖掘结直肠癌“炎癌转化”过程的关键基因，结合机器学习和神经网络模型的算法优势进行基因筛选，验证关键基因作为预测指标的可行性及免疫相关性，进而预测潜在防治中药。**方法**　利用基因表达数据库（Gene Expression Omnibus，GEO）2 个 GSE66407 和 GSE166925 数据集，筛选由正常结肠到炎症结肠（溃疡性结肠炎），最后进展到肠癌的过程中出现的差异基因；通过构建蛋白质-蛋白质相互作用网络（protein-protein interaction，PPI）进一步筛选核心基因；运用梯度提升机、随机森林和决策树 3 种分类机器学习算法进行结局的预测学习，将 3 种机器学习算法筛选出的核心基因取交集，构建深度学习框架下的人工神经网络模型，对模型预测准确性进行内部及外部验证；实时定量聚合酶链反应（real-time quantitative polymerase chain reaction，RT-PCR）联合人蛋白质图谱数据库（Human Protein Atlas，HPA）的免疫组化数据分析关键基因在初发结直肠癌患者癌组织及正常组织中的表达；运用单样本基因集富集分析（single sample gene set enrichment analysis，ssGSEA）算法分析核心基因与免疫细胞之间的关联；通过生存分析筛选对结直肠癌具有预后意义的基因；最后通过核心基因进行相关中药的预测分析。**结果**　在结直肠炎癌转化全过程中，共鉴定出 152 个共同的差异基因。这些基因参与肿瘤免疫反应、炎症反应等生物过程；经由 3 种机器学习算法的筛选，最终确定了 8 个核心基因，包括组织金属蛋白酶抑制剂 1（tissue inhibitor of metalloproteinases 1，*TIMP1*）、基质金属蛋白酶 1（matrix metalloproteinase 1，*MMP1*）、趋化因子 CXC 配体 13（C-X-C motif chemokine ligand 13，*CXCL13*）、普列克底物蛋白（Pleckstrin，*PLEK*）、颗粒酶 B（granzyme B，*GZMB*）、*CXCL5*、*CXCL3* 和 *CXCL8*。在深度学习框架下的神经网络模型中，训练集中对正常组和肿瘤组的预测准确率分别为 86.3%、92.3%；外部验证集中对正常组和肿瘤组的预测准确率分别为 80.0%、80.6%。8 个核心基因同时涉及多种免疫浸润过程；RT-PCR 结合 HPA 数据库分析显示 *TIMP1*、*MMP1*、*GZMB*、*CXCL5*、*CXCL8*、*CXCL3* 在结直肠癌患者癌组织中表达显著升高（$P<0.05$），CXCL13 在癌组织中表达显著降低（$P<0.05$），*PLEK* 在癌组织中表达降低但差异不显著；*TIMP1*、*CXCL13* 和 *CXCL8* 对患者的整体生存具有显著影响。对实验验证有显著差异的 7 个核心基因进行防治中药预测分析，结果显示预测中药以清热药和补虚药为主，四气以寒、温和平性为主，五味以苦、辛、甘味比例最大。**结论**　*TIMP1*、*MMP1*、*CXCL13*、*GZMB*、*CXCL5*、*CXCL3* 和 *CXCL8* 可以作为预测结直肠癌“炎-癌”转化过程的关键分子，对这些分子进行中药的早期干预，可能对结直肠癌的早期诊疗具有重要意义。

关键词：炎癌转化；结直肠癌；机器学习；神经网络模型；免疫浸润；清热药；补虚药

中图分类号：R285；TP391　　**文献标志码：**A　　**文章编号：**0253 - 2670(2023)19 - 6386 - 14

DOI: 10.7501/j.issn.0253-2670.2023.19.020

Identification of key genes in “inflammation-cancer transformation” process of colorectal cancer and prediction of traditional Chinese medicines for prevention and treatment through machine learning and neural network models

LIU Yang[1], CAO Sai-ya[1], FENG Yue-jiao[2], WANG Jie[1, 2], CHEN Teng[1]

1. Department of General Surgery, Putuo Hospital Affiliated to Shanghai University of Traditional Chinese Medicine (Shanghai Putuo District Central Hospital), Shanghai 200062, China
2. Department of General Surgery, Clinical College of Putuo Clinical Center, Anhui Medical University, Shanghai 200062, China

收稿日期：2023-07-12

基金项目：国家自然科学基金面上项目（81973625）；国家科技部重点研发计划（2019YFC1316000）；上海市普陀区卫生健康系统临床特色专病建设项目（2019tszb01）；上海市中医医院第三届中医传承人才项目（2023zycc02）

作者简介：刘　洋（1994—），男，在读博士研究生，研究方向为中西医结合防治消化道恶性肿瘤。E-mail: lyang4973@163.com

***通信作者：**陈　腾（1961—），男，主任医师，教授。E-mail: chenteng3387@126.com

Abstract: Objective To mine crucial genes involved in the "inflammation-cancer transformation" process of colorectal cancer using public databases, screen genes combined the algorithm advantages of machine learning and neural network models, verify the feasibility and immune correlation of key genes as predictive indicators, and then predict the potential traditional Chinese medicine (TCM) for prevention and treatment. **Methods** Two data sets GSE66407 and GSE166925 from Gene Expression Omnibus (GEO) were used to screen the differential genes that appeared in the progression from normal colon to inflammatory colon (ulcerative colitis) and finally to intestinal cancer. The core genes were further screened by constructing protein-protein interaction network (PPI). Three classification machine learning algorithms, gradient elevator, random forest and decision tree, were used to predict the outcome. The core genes screened by the three machine learning algorithms were intersected to build an artificial neural network model under the deep learning framework, and the prediction accuracy of the model was verified internally and externally. The real-time quantitative polymerase chain reaction (RT-PCR) combined with immunohistochemical data of Human Protein Atlas (HPA) were used to analyze the expression of key genes in cancer tissues and normal tissues of patients with primary colorectal cancer. The association between core genes and immune cells was analyzed by single sample gene set enrichment analysis (ssGSEA). The genes with prognostic significance for colorectal cancer were screened by survival analysis. Finally, the prediction analysis of related TCM was carried out through core genes. **Results** A total of 152 common differential genes were identified in "inflammation-cancer transformation" process of colorectal cancer. These genes were involved in biological processes such as tumor immune response and inflammatory response. Through the screening of the three kinds of machine learning algorithms, eight core genes were determine ultimately, including tissue inhibitor of metalloproteinases 1 (*TIMP1*), matrix metalloproteinase 1 (*MMP1*), C-X-C motif chemokine ligand 13 (*CXCL13*), Pleckstrin (*PLEK*), granzyme B (GZMB), *CXCL5*, *CXCL3* and *CXCL8*. In the neural network model under the framework of deep learning, the prediction accuracy of the normal group and the tumor group were 86.3% and 92.3% in training set, respectively. The prediction accuracy of the normal group and the tumor group was 80.0% and 80.6% in validation set, respectively. A total of eight core genes involved in multiple immune infiltration processes simultaneously; RT-PCR combined with HPA database analysis showed that the expression of *TIMP1*, *MMP1*, *GZMB*, *CXCL5*, *CXCL8* and *CXCL3* in cancer tissues of colorectal cancer patients was significantly increased ($P < 0.05$), while the expression of *CXCL13* was significantly decreased ($P < 0.05$). PLEK expression decreased in cancer tissues but the difference was not significant. *TIMP1*, *CXCL13*, and *CXCL8* had a significant impact on overall survival. The prediction analysis of seven core genes with significant differences in experimental verification of TCM showed that the predicted TCM mainly consisted of heat-clearing and deficient-tonifying medicines, with cold, warm and peaceful in nature, and bitter, pungent and sweet in flavours. **Conclusion** *TIMP1*, *MMP1*, *CXCL13*, *GZMB*, *CXCL5*, *CXCL3* and *CXCL8* could be used as key molecules to predict the "inflammation-cancer transformation" process of colorectal cancer. The early intervention of TCM on these molecules may be of great significance for the early diagnosis and treatment of colorectal cancer.

Key words: inflammation-cancer transformation; colorectal cancer; machine learning; neural network model; immune infiltration; heat-clearing medicine; deficient-tonifying medicine

结直肠癌是全球发病率和死亡率较高的恶性肿瘤之一，尽管诊断和治疗技术的进步已经显著提高了结直肠癌患者的生存率，但对于疾病的早期诊断和有效预测仍然存在挑战。尤其是炎症和癌症之间的紧密关系引发了科学家对于炎癌转化机制的深入研究[1]。炎症性肠病（inflammatory bowel disease，IBD）是一组包括克罗恩病（Crohn's disease，CD）和溃疡性结肠炎（ulcerative colitis，UC）在内的慢性炎症性肠道疾病[2]。近年来，越来越多的研究表明，炎症性肠病与结直肠癌发生的风险密切相关，其中炎症环境对结直肠癌的“炎-癌”转化过程具有重要影响。在这个过程中，炎症反应激活了一系列的信号转导途径，影响了细胞的生长、分化和存活，最终导致了肿瘤的形成[3]。然而，这些过程中的关键驱动因素和具体机制仍不完全清楚。因此，通过对结直肠癌炎症-肿瘤转化过程的深入研究，了解其分子基础，对于发现新的预防和治疗策略具有重要的实际意义。

在中医药的理论体系中，“未病先防”是一种重要的治疗理念，其主张在疾病尚未发生之前采取预防措施，旨在提高人体自身的抵抗能力，防止疾病的发生[4]。近年来，西医领域对疾病的理解也逐渐转向这一思路，尤其是在癌症研究中，防控癌症早期发生的研究得到了广泛关注。结直肠癌的“炎-癌”转化过程，即长期存在的慢性炎症状态可能诱导结直肠正常细胞向恶性转化，进而形成肿瘤。对这一过程的深入理解，对于防治结直肠癌具有重要的实际意义。

尽管近年来对于这种“炎-癌”转化的研究取得了一些进展，但其具体的分子机制和关键调控因子仍不完全清楚。为了深入理解这一过程，本研究采用机器学习算法对大规模的转录组数据进行分析，以期找到在结直肠癌“炎-癌”转化过程中的关键分子，并对这些分子在疾病进程中的作用进行深入探讨，最后通过核心基因对关联中药进行预测。本研究希望为结直肠癌的临床诊断和中医药的治疗提供潜在的生物标志物和药物靶点。

1 材料与方法

1.1 数据获取

本研究所有数据均来源于公共数据库，利用基因表达数据库（Gene Expression Omnibus，GEO）的 2 个数据集筛选差异基因，GSE66407 用于溃疡性结肠炎炎症期患者与正常非炎症患者差异基因筛选；GSE166925 用于溃疡性结肠炎患者与结直肠癌患者间的差异基因筛选。肿瘤基因图谱（Cancer Genome Atlas，TCGA）数据库中的结直肠癌数据集用于机器学习及神经网络模型中的基因筛选，因正常样本数目与肿瘤样本数差异较大，本研究引入基因-组织表达（Genotype-Tissue Expression，GTEx）数据库补充正常患者样本，使用 R 语言中的 sva 包[5]去除批次效应用于后续分析；GSE44076 作为外在验证集验证神经网络模型的准确性及后续的免疫浸润分析。

1.2 差异基因筛选

Limma 包[6]及 DEseq2 包[7]分别用于 GSE66407 及 GSE166925 中差异基因的筛选，筛选阈值设定为 $\log_2FC>1$，$P<0.05$，ggVennDiagram[8]、ggplot2[9]包用于绘制 Venn 图显示正常结肠-溃疡性结肠炎-结直肠癌转化过程中的差异基因，以 Venn 图形式展现“炎-癌”转化过程中关键分子的交集情况。

1.3 交集基因富集分析及关键基因蛋白质-蛋白质相互作用网络（protein-protein interaction，PPI）网络构建

为了深入探究在结直肠癌炎症至癌变转化过程中所涉及的细胞通路和分子过程，对在这个过程中的交集差异基因进行基因本体（gene ontology，GO）和京都基因与基因组百科全书（Kyoto encyclopedia of genes and genomes，KEGG）富集分析，R 语言中的 clusterProfiler 软件包用于基因的富集分析。

在 GO 分析中，从生物过程（biological process，BP）、分子功能（molecular function，MF）以及细胞组成（cellular component，CC）3 个层面展现富集情况。KEGG 分析是为了了解差异基因在各类生物通路中的作用和位置，尤其是在癌症相关通路中的角色。以上 2 种分析都采用了超几何检验（hypergeometric test）来确定某一 GO 功能或 KEGG 通路的富集是否显著，并计算其显著性水平（*P* 值）。富集分析过程由 R 语言 clusterProfiler[10]包实现。

在差异表达基因导入 STRING（https://string-db.org/）数据库进行 PPI 分析，将物种设置为“homo sapiens”（人类），并选择相互作用得分大于 0.4 的基因。随后，使用 Cytoscape 进行 PPI 网络的可视化。由于网络过于复杂，不便直接展示，因此使用 CytoHubba 插件[11]来筛选出 PPI 网络中度（degree）值前 30 的核心基因用于后续筛选过程。

1.4 机器学习算法筛选结直肠癌炎癌转化过程的生物标志物

机器学习是基于数据构建的计算模型，具有良好的可解释性。在获取结直肠癌炎癌转化 2 个阶段的核心基因后，采用机器学习算法来进一步进行基因筛选。选用 3 种不同的机器学习算法进行分析，分别是梯度提升机（gradient boosting machine，GBM）算法、随机森林（random forest，RF）算法和决策树（decision tree，DT）算法，设定“feature importance”（特征重要性）大于 2 用于筛选机器学习模型中的重要基因。

1.5 神经网络模型预测结直肠癌的发生

使用神经网络模型来构建预测结直肠癌发生的模型。神经网络是一种模仿人脑神经元工作机制的复杂模式识别模型，其可以从输入数据中学习并做出预测。在本研究中使用了前馈神经网络（feedforward neural network，FNN），这是一种常见的神经网络模型，其中信息的传递只能从输入层向前到隐藏层，再向前到输出层，不存在反向的信息传递。

使用 TCGA 联合 GTEx 的数据集作为神经网络模型训练集，包含 453 个肿瘤样本，349 个正常样本，输出模型的准确性及受试者工作特征（receiver operator characteristic，ROC）曲线；利用 GSE44076（包含 98 个结直肠癌肿瘤样本及 50 个正常样本）作为外在验证集进行神经网络模型的验证。借助于 neuralnet R 语言包，搭建了人工神经网络模型。首先执行数据预处理，目标是实现数据的规范化。接着，选用范围在 [0，1] 的 min-max 方法，将数据进行分离缩放，这是在训练神经网络前的重要步骤。

随后，对数据的最大、最小值进行标准化处理，并将神经网络的隐藏层设定为 5 层。

1.6 关键基因与肿瘤细胞中的免疫浸润关联性分析

为了深入了解筛选出的预测标志物在肿瘤细胞中的免疫浸润情况，利用 GSE44076 的表达数据进行单样本基因集富集分析（single-sample gene set enrichment analysis，ssGSEA），将核心基因的表达模式与已知的免疫细胞基因组特征进行比较，来评估核心基因在肿瘤免疫细胞中的富集程度，ssGSEA 通过计算每个样本中基因集的富集得分来量化基因集在样本中的表达模式。

1.7 关键基因在结直肠癌与正常组织中的表达差异分析

从上海中医药大学附属普陀医院结直肠癌样本库中选取 12 组初发结直肠癌患者的癌组织与正常组织，使用实时定量聚合酶链反应（real-time quantitative polymerase chain reaction，RT-PCR）检测关键基因在癌组织与正常组织中的表达差异。采用 Trizol 法提取组织中的总 RNA，取 500 ng RNA 逆转录为 cDNA（Takara，RR036A）；使用 TB green（Takara，RR420）染料按照 10 μL 体系进行 45 次循环的扩增，引物序列如表 1 所示，以 3-磷酸甘油醛脱氢酶（glyceraldehyde-3-phosphate dehydrogenase，GAPDH）为内参，按 $2^{-\Delta\Delta Ct}$ 法分析。从人蛋白质图谱数据库（Human Protein Atlas，HPA）数据库中寻找核心基因在癌组织与正常组织中的免疫组化表达差异。

1.8 关键基因生存分析

基于 GEPIA 数据库[12]（http://gepia.cancer-pku.cn/）进行关键基因的总生存期（overall survival，OS）的 KM 曲线分析，默认分组截断值为中位值。

1.9 靶向中药预测

基于 Coremine Medical 数据库（https://www.coremine.com/medical/）预测调控核心基因的潜在中药，筛选 $P<0.01$ 的中药纳入分析，搜集并整理中药的四气五味及功效主治信息进行统计和分析。

1.10 统计学方法

本研究利用 R 语言中的 limma、survival、DESeq2 以及 ggplot2 包等进行数据处理、统计分析和结果可视化；利用 GraphPad 9.5.1 进行 RT-PCR 的统计分析，两组样本的数据符合正态分布且方差齐性时，组间比较使用独立样本 *t* 检验；两组样本的数据不符合正态分布或者方差不齐时，则使用非参数检验。

表 1 引物序列

Table 1 Primer sequences of genes

基因	序列（5’-3’）
TIMP1	上游：CTTCTGCAATTCCGACCTCGT
	下游：ACGCTGGTATAAGGTGGTCTG
PLEK	上游：AAGAAGGGGAGCGTGTTCAAT
	下游：TCAGCGGGATCATTCCTTTGG
MMP1	上游：AAAATTACACGCCAGATTTGCC
	下游：GGTGTGACATTACTCCAGAGTTG
GZMB	上游：CCCTGGGAAAACACTCACACA
	下游：GCACAACTCAATGGTACTGTCG
CXCL8	上游：TTTTGCCAAGGAGTGCTAAAGA
	下游：AACCCTCTGCACCCAGTTTTC
CXCL5	上游：AGCTGCGTTGCGTTTGTTTAC
	下游：TGGCGAACACTTGCAGATTAC
CXCL3	上游：CGCCCAAACCGAAGTCATAG
	下游：GCTCCCCTTGTTCAGTATCTTTT
CXCL13	上游：GCTTGAGGTGTAGATGTGTCC
	下游：CCCACGGGGCAAGATTTGAA
GAPDH	上游：GGAGCGAGATCCCTCCAAAAT
	下游：GGCTGTTGTCATACTTCTCATGG

2 结果

2.1 数据预处理展示

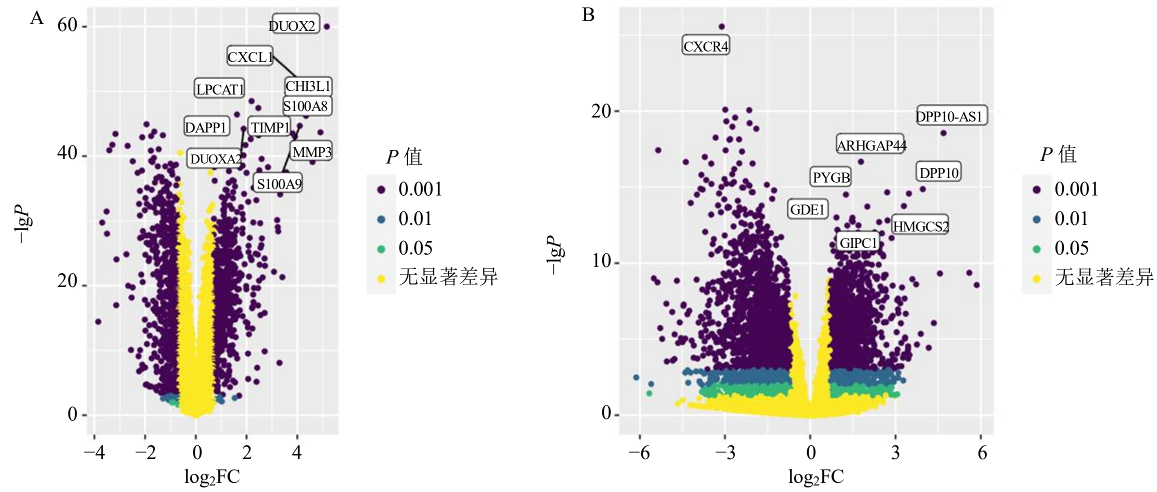
相比于正常非炎症患者，溃疡性结肠炎炎症期患者的肠道组织中共有 401 个基因上调，345 个基因下调（设定 $\log_2FC>1$，$P<0.05$），见图 1。相比于溃疡性结直肠癌患者，在结直肠癌患者肠道肿瘤组织中，共有 212 个基因上调，600 个基因下调（设定 $\log_2FC>1$，$P<0.05$），满足设定条件的基因展示在火山图上。

2.2 差异基因筛选

在 2 个炎癌转化阶段的差异基因中，取交集后得到共同的 152 个差异基因，如图 2 所示。在这些基因中，发现了 4 个在 2 个阶段中均被上调的基因，分别是前列腺癌易感性候选基因 1（prostate cancer susceptibility candidate 1，*PRAC1*）、胰岛素样肽 5（insulin-like peptide 5，*INSL5*）、细胞表面相关黏蛋白 12（mucin 12，*MUC12*）和癌胚抗原相关细胞黏附分子 7（carcinoembryonic antigen-related cell adhesion molecule 7，*CEACAM7*），这 4 个基因亦被纳入后续筛选过程。

2.3 交集差异基因富集分析

富集分析结果（图 3）显示，在生物过程中，中性粒细胞趋化与迁移、白细胞趋化与迁移、粒细胞



A-正常非炎症患者 *vs* 溃疡性结肠炎炎症期患者　B-溃疡性结肠炎患者 *vs* 结直肠癌患者

A-normal non-inflammatory patients *vs* patients with ulcerative colitis inflammation phase　B-ulcerative colitis patients *vs* colorectal cancer patients

图 1　差异表达基因火山图

Fig. 1　Volcano plots of differential expressed genes

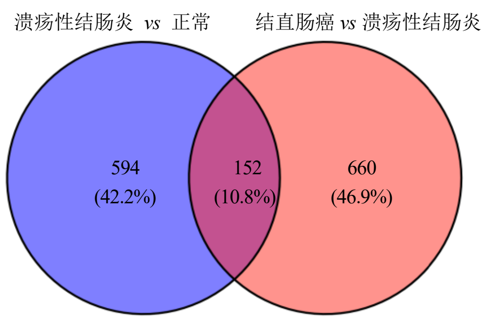


图 2　炎癌转化 2 个阶段差异基因 Venn 图

Fig. 2　Venn diagram of differential expressed genes across two stages of inflammation-cancer transformation

趋化与迁移显著富集，这些富集的生物过程主要涉及白细胞的迁移和化学趋性，暗示了免疫细胞可能在炎癌转化过程中发挥了中心作用；细胞成分的富集分析揭示了含胶原的细胞外基质、分泌颗粒腔、囊泡腔、细胞突起膜在炎癌转化过程中可能具有重要作用，可能与炎癌转化的细胞行为改变有关；在分子功能方面，细胞因子活性、趋化因子受体结合、多糖结合、免疫受体活性和细胞因子结合显著富集，这些分子功能的富集揭示了免疫反应和细胞信号转导在炎癌转化过程中可能发挥关键作用。KEGG 通

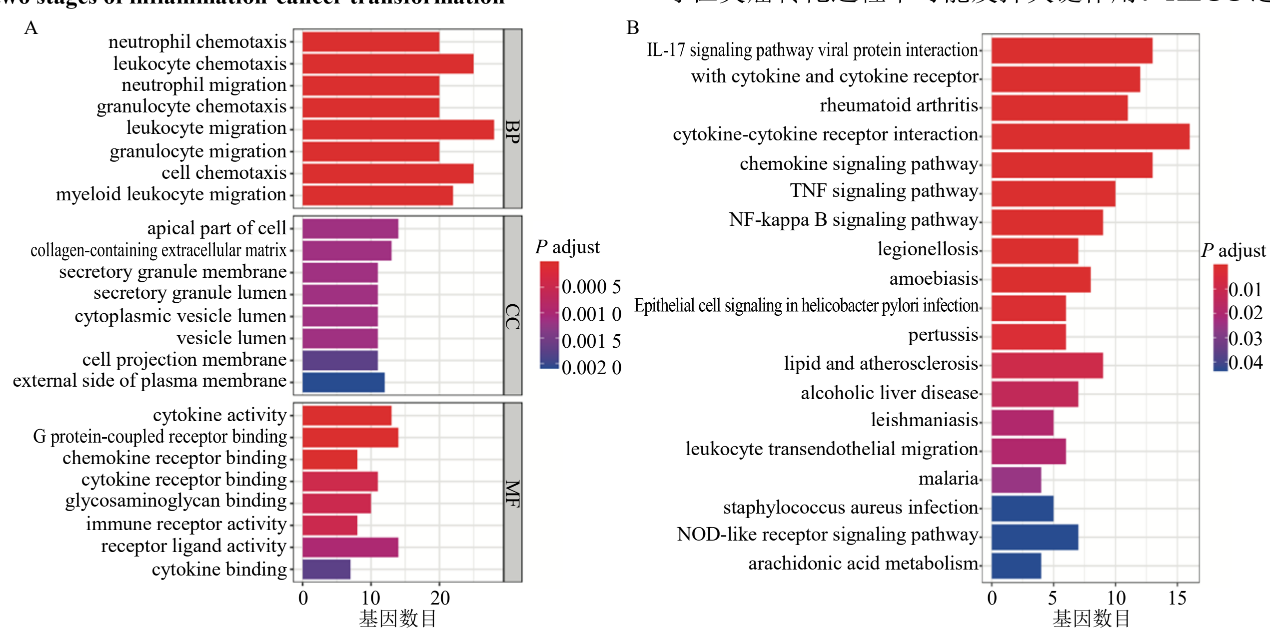


图 3　结直肠癌炎癌转化过程差异基因 GO 功能 (A) 和 KEGG 通路 (B) 富集分析

Fig. 3　GO function (A) and KEGG pathway (B) enrichment analysis of differential expressed genes in "inflammation-cancer transformation" process of colorectal cancer

路富集分析揭示了白细胞介素-17（interleukin 17，IL-17）信号通路、病毒与细胞因子和细胞因子受体的相互作用、类风湿性关节炎、细胞因子-细胞因子受体相互作用、趋化因子信号通路、肿瘤坏死因子（tumor necrosis factor，TNF）信号通路、核因子-κB（nuclear factor-κB，NF-κB）信号通路等在炎癌转化过程中可能具有重要作用，这些富集的通路可能影响炎癌转化过程，为未来的研究提供潜在的疾病标志物和治疗靶点。

2.4 差异基因 PPI 网络构建及关键（hub）基因筛选

通过对差异基因进行 PPI 网络分析，并使用 Cytoscape 的 CytoHubba 插件根据网络 degree 值进行筛选，鉴定出了 30 个可能在结直肠癌炎癌转化过程中发挥重要作用的 hub 基因，核心基因 PPI 网络图见图 4，图中圆圈的大小和颜色深度代表该基因在网络中的 degree 值，这些基因大致可分为 4 种类别：（1）细胞因子及其受体类基因，如蛋白质酪氨酸磷酸酶受体 C 型（protein tyrosine phosphatase receptor type C，*PTPRC*）、白细胞介素 1 beta（interleukin 1 beta，*IL1B*）、趋化因子 CXC 配体 8（C-X-C motif chemokine ligand 8，*CXCL8*）、*CXCL1*、前列腺素内过氧化物合酶 2（prostaglandin-endoperoxide synthase 2，PTGS2）、CXC 趋化因子受体 4（C-X-C chemokine receptor 4，CXCR4）、趋化因子受体 7（C-C motif chemokine receptor 7，CCR7）、

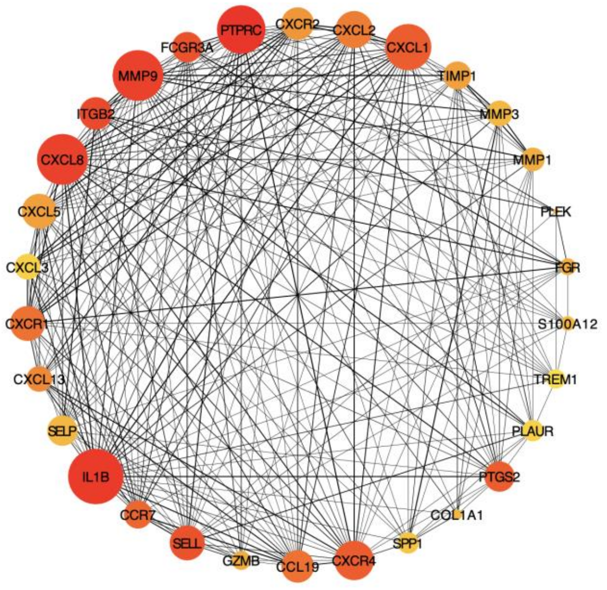


图 4 核心差异基因 PPI 网络

Fig. 4 PPI network diagram of core differential expressed genes

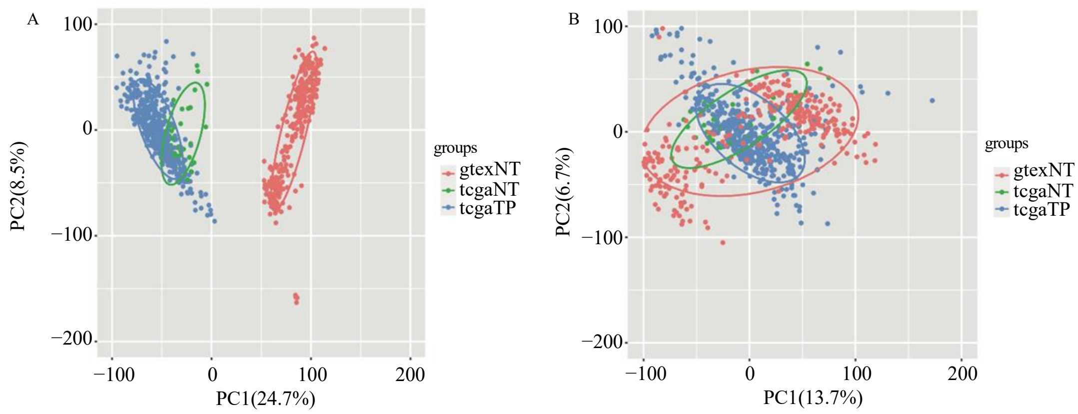
CXCR1、趋化因子配体 19（C-C motif chemokine ligand 19，*CCL19*）、*CXCL2*、*CXCL13*、*CXCR2*、*CXCL5* 和 *CXCL3*，这些基因在炎症反应和免疫调控中起关键作用，尤其是对白细胞的趋化和活化有重要影响；（2）蛋白酶及其抑制因子类基因，如基质金属蛋白酶 9（matrix metalloproteinase 9，*MMP9*）、*MMP1*、组织金属蛋白酶抑制剂 1（tissue inhibitor of metalloproteinases 1，*TIMP1*）和 *MMP3*，它们主要参与细胞外基质的重塑，与肿瘤的侵袭性增强可能有密切关系；（3）细胞表面蛋白类基因，如整合素亚基 beta 2（integrin subunit beta 2，*ITGB2*）、选择素 L（selectin L，*SELL*）、免疫球蛋白 γ Fc 区域受体 III-A（immunoglobulin γ Fc region receptor III-A，*FCGR3A*）和选择素 P（selectin P，*SELP*），这些基因涉及白细胞的迁移和与肿瘤微环境的交互；（4）其他类别的基因，如普列克底物蛋白（Pleckstrin，*PLEK*）、原癌基因 C-Fgr（proto-oncogene C-Fgr，*FGR*）、颗粒酶 B（granzyme B，*GZMB*）、胶原蛋白 α-1(I) 链［collagen alpha-1(I) chain，*COL1A1*］、中性粒细胞 S100 蛋白（neutrophil S100 protein，*S100A12*）、分泌型磷蛋白 1（secreted phosphoprotein 1，*SPP1*）、纤溶酶原激活尿激酶受体（plasminogen activator urokinase receptor，*PLAUR*）和骨髓细胞触发受体 1（triggering receptor expressed on myeloid cells 1，*TREM1*），它们可能涉及到多种生物过程，包括免疫调控、细胞迁移和增殖、细胞外基质组织以及信号传导等。

2.5 机器学习算法筛选预测基因

利用机器学习算法对 30 个核心的 hub 基因进行进一步筛选，首先进行 GTEx 数据集和 TCGA 数据库的合并，去除批次效应，主成分分析（principal component analysis，PCA）图见图 5，共纳入 453 个肿瘤样本，349 个正常样本，获取核心 30 个 hub 基因的表达量后进行分类机器学习算法分析，利用 3 种机器学习算法 GBM、RF、DT 来筛选出对预测结局有意义的目标基因，3 种机器学习算法筛选出的基因及交集情况见图 6 和表 2，在这 3 种机器学习算法中，有 8 个共同的基因被筛选出来，分别是 *TIMP1*、*MMP1*、*CXCL13*、*PLEK*、*GZMB*、*CXCL5*、*CXCL3*、*CXCL8*。

2.6 神经网络模型构建预测结直肠癌发生准确度

构建的神经网络模型（图 7-a）预测结局的准确度，在训练集中对正常组预测准确率为 86.3%，肿

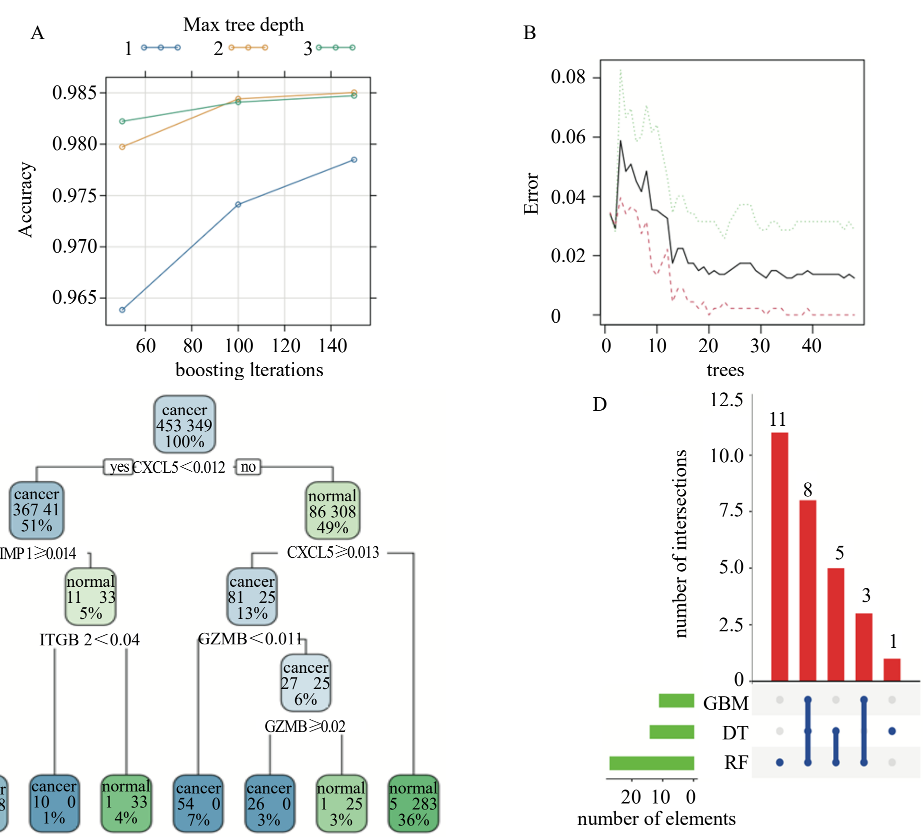


A-去批次前　B-去批次后　NT-正常组织　TP-初发肿瘤

A-before batch correction　B-after batch correction　NT-normal tissue　TP-primary tumors

图 5　TCGA 数据库与 GTEx 数据库合并后的批次效应图

Fig. 5　Batch effect diagram after merging TCGA database and GTEx database



A-GBM 算法　B-RF 算法（点虚线为总误差，实线为预测肿瘤误差，虚线为预测正常误差）　C-DT 算法　D-3 种算法基因交叉情况

A-GBM algorithm　B-RF algorithm (dot-dashed line represent total error　full line represent error in predicting tumors　dashed line represent error in predicting normal tissue)　C-DT algorithm　D-intersection of genes across three algorithms

图 6　3 种机器学习算法模型

Fig. 6　Three machine learning algorithm models

瘤组预测准确率为 92.3%。ROC 曲线的曲线下面积（area under curve，AUC）值可达 0.956，见图 7-b；外在验证集中正常组预测准确率为 80%，肿瘤组预测准确率为 80.6%，ROC 曲线的 AUC 值可达 0.876，

表 2 3 种机器学习算法核心基因及交集基因

Table 2 Core genes identified by three machine learning algorithms and intersecting genes

机器学习算法	关键基因
GBM 算法	*CXCL5*、*GZMB*、*PLEK*、*TIMP1*、*CXCL8*、*CXCL13*、*TREM1*、*MMP1*、*SPP1*、*CXCL3*、*SELP*
随机森林算法	*CXCL5*、*MMP1*、*PLEK*、*GZMB*、*CXCL8*、*CXCL13*、*TIMP1*、*MMP9*、*TREM1*、*CXCL2*、*IL1B*、*SELL*、*CXCR1*、*CXCL3*、*CXCL1*、*COL1A1*、*MMP3*、*PLAUR*、*PTPRC*、*CCL19*、*FCGR3A*、*FGR*、*PTGS2*、*SELP*、*SPP1*、*CCR7*、*CXCR2*
决策树算法	*CXCL5*、*CXCL8*、*MMP1*、*CXCL13*、*PLEK*、*GZMB*、*TIMP1*、*CXCL3*、*COL1A1*、*CXCL1*、*CCL19*、*ITGB2*、*FGR*、*MMP9*
交集基因	*TIMP1*、*MMP1*、*CXCL13*、*PLEK*、*GZMB*、*CXCL5*、*CXCL3*、*CXCL8*

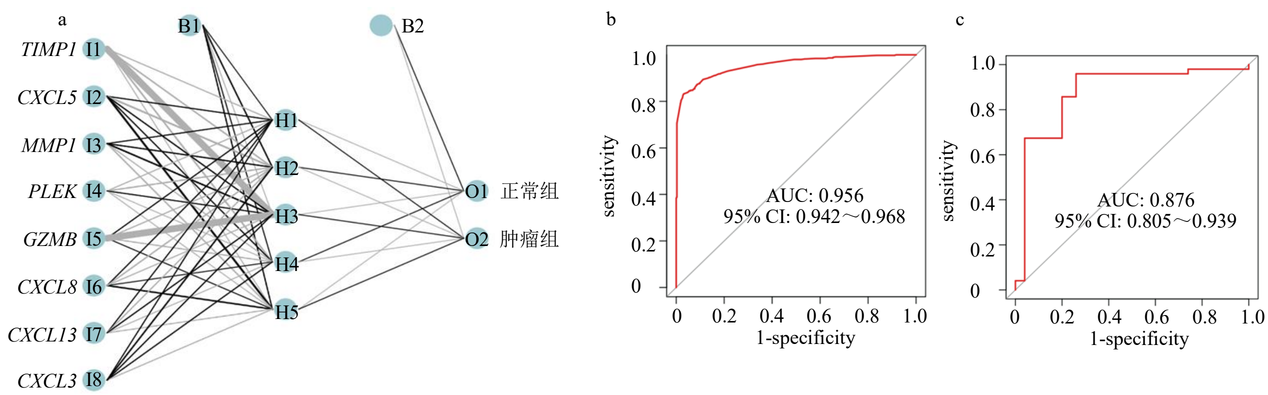


图 7 神经网络模型 (a)、训练集 ROC 曲线 (b) 和验证集 ROC 曲线 (c)

Fig. 7 Neural network model (a), ROC curve for training set (b) and ROC curve for validation set (c)

ROC 曲线见图 7-c。

2.7 核心基因免疫相关性分析

通过 ssGSEA 分析发现 8 个核心基因在结直肠癌中与免疫细胞类型呈现显著的相关性，表明这些基因可能在调控肿瘤的免疫浸润中起到关键作用，并可能涉及结直肠癌的免疫调控过程。*TIMP1*、*PLEK*、*CXCL13* 在多数免疫细胞类型中表现出强烈的相关性，可能在结直肠癌的炎症至癌症转化过程中起到重要的作用。这进一步提示 *TIMP1*、*PLEK*、*CXCL13* 在结直肠癌的免疫调控及免疫浸润调节中可能具有显著的功能。同时，*MMP1*、*GZMB*、*CXCL8*、*CXCL5*、*CXCL3* 在部分免疫细胞类型中与结直肠癌呈现相关性，尤其在活化的 CD4 T 细胞、自然杀伤细胞以及先天免疫细胞中，这些基因的表达与结直肠癌的相关性尤为突出，见图 8。

2.8 核心基因在癌组织和正常组织中的表达

RT-PCR 结果显示，*TIMP1*、*MMP1*、*GZMB*、*CXCL3*、*CXCL5*、*CXCL8* 在肿瘤组织中表达显著升高（$P<0.05$），*CXCL13* 在肿瘤组织中表达显著降低（$P<0.05$），*PLEK* 在癌组织和正常组织中表达无显著差异，见图 9；HPA 数据库中匹配到 6 个核心基因的免疫组化数据，结果显示，*TIMP1*、*GZMB*、*CXCL5*、*CXCL8*、*CXCL13* 在肿瘤组织中的表达升高，*CXCL13* 在正常组织中表达升高，*PLEK* 在两组表达中无显著差异，见图 10。

2.9 核心基因生存分析

通过对 TCGA 数据库进行生存分析，发现在筛选出的 8 个核心基因中，有 3 个核心基因（*TIMP1*、*CXCL13* 和 *CXCL8*）的 $P<0.05$，具有临床意义。KM 曲线生存分析图见图 11，KM 曲线显示高表达 *TIMP1* 与结直肠癌患者的生存状况呈不良影响关联，而低表达的 *CXCL13*、*CXCL8* 与患者的生存状况也呈不良影响关联。

2.10 核心中药预测

基于筛选及验证结果，将 *TIMP1*、*MMP1*、*GZMB*、*CXCL3*、*CXCL5*、*CXCL8*、*CXCL13* 7 个核心基因进行中药预测，*CXCL3* 在限定条件下未匹配到目标中药，最终基因与中药的预测网络图见图 12，四气五味及功效主治统计结果见图 13。结果显示预测中药以清热药和补虚药为主，四气以寒、温

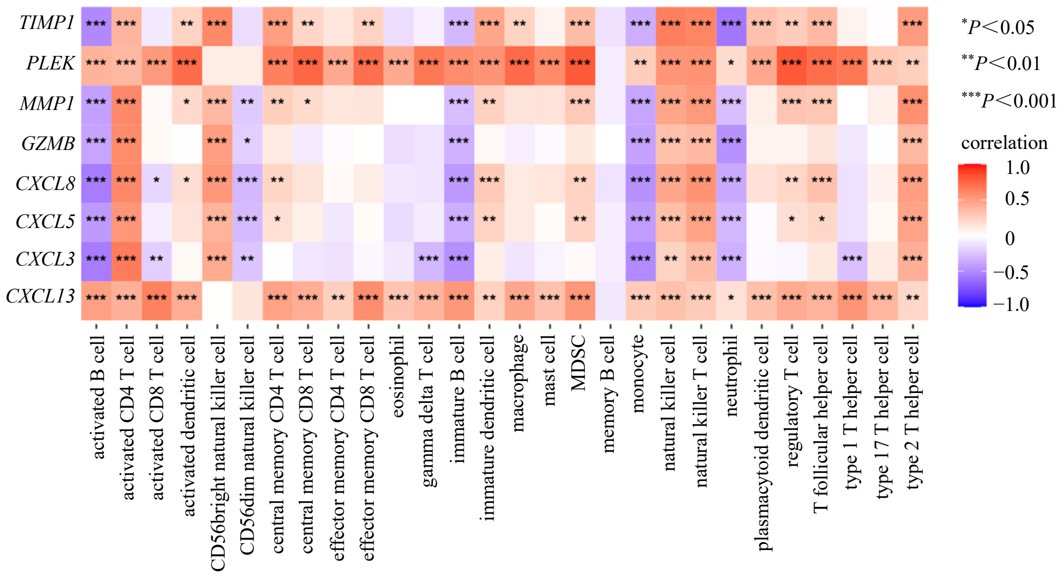
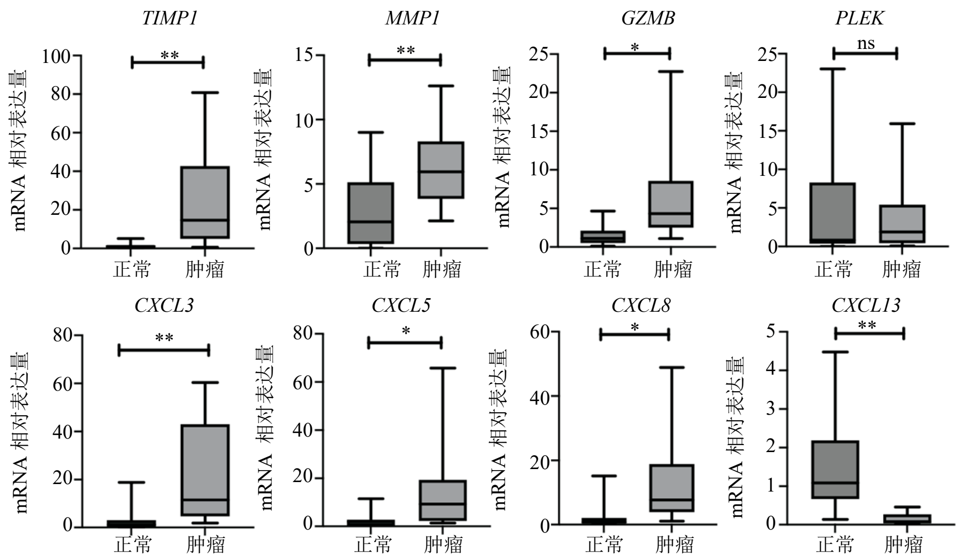


图 8　核心基因的 ssGSEA

Fig. 8　ssGSEA of core genes



*P＜0.05　**P＜0.01　ns-无显著性

*P < 0.05　**P < 0.01　ns-no significant difference

图 9　核心基因在癌组织与正常组织中的表达差异

Fig. 9　Differential expression of core genes in cancerous and normal tissues

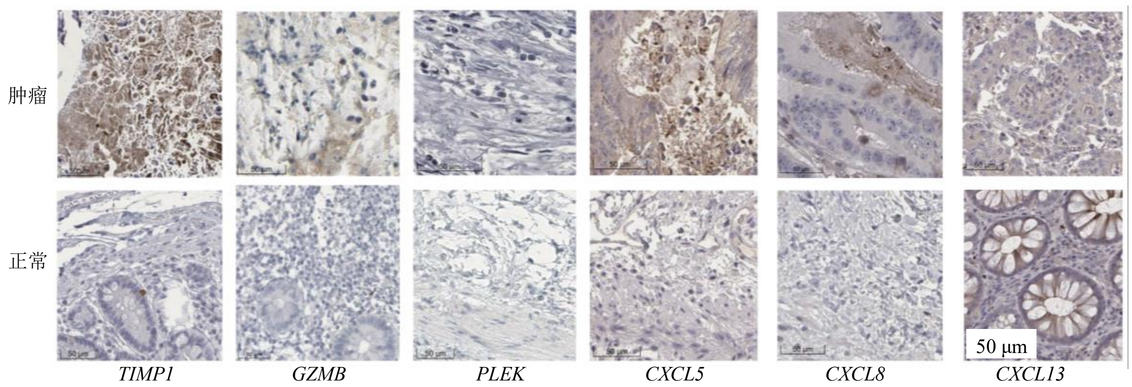


图 10　HPA 数据库中 6 个核心基因的表达差异

Fig. 10　Differential expression of six core genes in HPA database

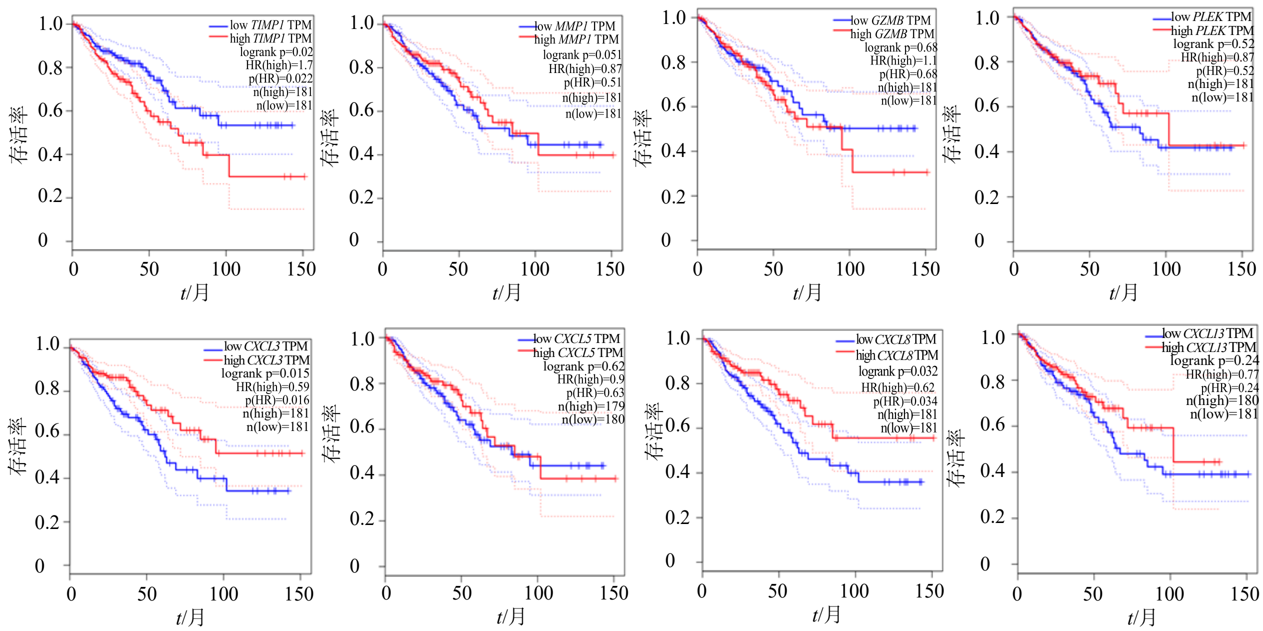


图 11 关键基因生存分析

Fig. 11 Survival analysis of key genes

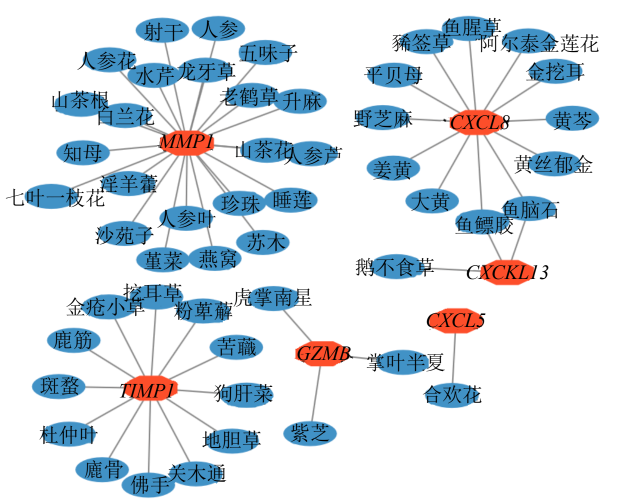


图 12 核心基因中药预测网络图

Fig. 12 Network diagram of traditional Chinese medicines prediction for core genes

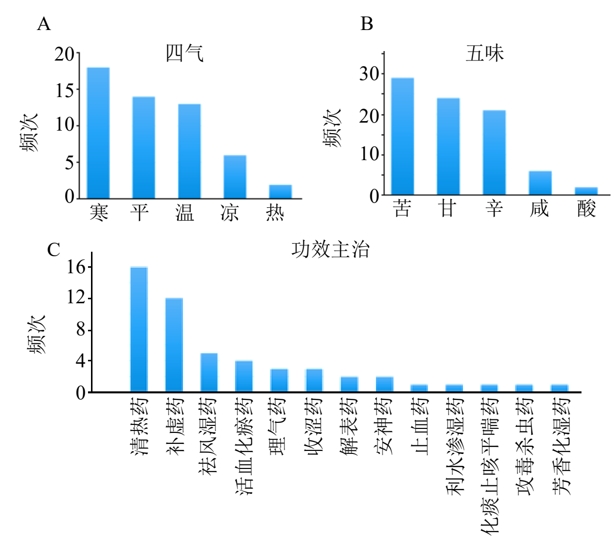


图 13 核心中药性味归经和功能主治统计

Fig. 13 Statistics on properties, flavors, meridian tropism, and main functions of core traditional Chinese medicines

和平性为主，五味以苦、辛、甘味比例最大。

3 讨论

3.1 炎症过程及通路在结直肠癌发生发展过程扮演重要的作用

结直肠癌是全球常见的恶性肿瘤，其发病机制与炎症过程密切相关[13]。在肠道肿瘤研究中，炎症环境被认为是促进肿瘤形成和发展的重要因素，这一过程被称为“炎癌转化”，同时，肿瘤细胞可以通过多种机制逃避免疫监视，从而使肿瘤得以持续发展[14]。结直肠癌的“炎癌转化”过程是一个极其复杂且微观的生物过程，本研究首先通过筛选结直肠癌炎癌转化 2 个阶段（正常结肠-炎性结肠，炎症结肠-结直肠癌）共同的差异基因，根据差异基因的富集分析发现炎癌转化阶段的诸多炎症通路和分子过程。

本研究的富集分析结果揭示了这一过程中可

能涉及的关键生物过程、细胞成分和分子功能，以及可能发挥重要作用的 KEGG 通路，从而加深了对结直肠癌"炎癌转化"过程的理解。GO 生物过程的富集分析结果揭示了白细胞，特别是中性粒细胞和粒细胞在"炎癌转化"过程中可能发挥了重要作用；这些细胞的趋化和迁移在这一过程中显著富集，表明免疫细胞在炎癌转化过程中扮演了中心角色，这一发现与近年来对免疫微环境在肿瘤发生发展中作用的认识一致，也为肿瘤免疫治疗策略的探索提供了新的思路。细胞成分的富集分析揭示了细胞行为，如细胞交互、分泌和细胞极性等在炎癌转化过程中的改变；在分子功能方面，细胞因子活性、趋化因子受体结合、细胞因子受体结合、受体配体活性和细胞因子结合在"炎癌转化"过程中可能发挥关键作用，这些分子功能的富集揭示了免疫反应和细胞信号转导在炎癌转化过程中可能发挥关键作用。KEGG 通路富集分析结果显示，在"炎癌转化"过程中发挥关键作用的通路包括 IL-17 信号通路、细胞因子-细胞因子受体相互作用、趋化因子信号通路、TNF 信号通路、NF-κB 信号通路等，些通路可能是结直肠癌炎癌转化过程的关键调控因子，也可能是未来研究的重要靶点。

3.2 *TIMP1*、*MMP1*、*CXCL13*、*GZMB*、*CXCL5*、*CXCL3* 及 *CXCL8* 预测结肠炎症过程的恶性转化

机器学习和神经网络模型都是目前人工智能的计算模型，用于处理和解释数据，从数据中学习和提取有用的模式，以用于预测、分类、决策等。本研究采用的 3 个机器学习模型 RF、DT、GBM 都可以很好地进行二分类任务，3 种方法各有优略，但本质都是基于数据的计算模型，具有良好的可解释性，因此本研究将其用于筛选区分癌症与正常样本的核心基因。关于神经网络模型，一般认为隐藏层≥2 的为深度学习网络，其利用人脑的神经元原理进行复杂的运算，相比之下，神经网络通常被视为"黑箱"，本研究利用这个"黑箱"作用，将 8 个核心基因作为神经网络的输入，结果作为输出，对 8 个核心基因预测结果的准确性进行了内部和外部 2 次验证，并利用常规机器学习算法的可解释性对核心基因进行进一步筛选，利用深度学习框架下的神经网络模型对核心基因的预测结果（癌症和正常样本）进行准确性的判断。

近年来，随着机器学习和神经网络模型在生物医学领域的广泛应用，为深入理解和预测炎癌转化过程提供了新的可能[15]。基于这些差异基因构建 PPI 网络筛选出 30 个核心基因，再使用机器学习算法，成功鉴定出了结直肠癌过程中的 8 个关键基因，这些基因与结直肠癌的炎癌转化过程密切相关，ssGSEA 分析显示这些基因与诸多免疫细胞相关，对后续研究具有一定指导意义。PPI 网络中富集到的 30 个核心基因可归类为细胞因子及其受体、蛋白酶及其抑制因子、细胞表面蛋白类，这 3 类基因的蛋白在肿瘤微环境中均起到非常重要的作用。

RT-PCR 结合免疫组化结果显示 *TIMP1*、*MMP1*、*GZMB*、*CXCL3*、*CXCL5*、*CXCL8* 在肿瘤组织中表达显著升高，*CXCL13* 在肿瘤组织中表达显著降低，意味着 7 个基因在预测结直肠癌炎癌转化的过程中可以作为潜在的作用靶点，这些基因在其他肿瘤中的炎症过程中亦有相关研究。

TIMP1 的水平在浸润肿瘤的淋巴细胞中显著增加，并与癌症进展相关[16]。*TIMP1* 在许多类型的癌症中表达增加，包括胃癌、胰腺癌、结直肠癌和乳腺癌[17-18]。一些研究发现，*TIMP1* 的水平与胃癌的分期以及对远处器官的转移有显著的相关性[19]；在胰腺癌中，*TIMP1* 表达水平较高的患者对治疗的反应较差，且更容易发生肝转移[20]。*MMP1* 编码降解细胞外基质金属蛋白酶，现有研究表明其与头颈癌[21]、肝癌[22]、肺癌[23]等肿瘤的发生发展相关，并参与肿瘤微环境的炎症过程[24]，在结直肠癌中，肿瘤相关巨噬细胞通过 *MMP1* 起到促进肿瘤发展的作用[25]，还与各种结肠炎、结肠腺瘤的发生发展密切相关[26]。*PLEK* 在肿瘤中的研究较少，有研究表明其与慢性牙周炎相关[27]。*GZMB* 是 1 种由 NK 细胞和 T 细胞产生的丝裂原，能够诱导肿瘤细胞凋亡[28]。

趋化因子是一类小分子蛋白，它们可以在细胞间产生化学梯度，从而引导细胞移动。在肿瘤微环境中，趋化因子有多种作用，包括促进炎症、调节免疫应答，以及直接影响肿瘤细胞的生长和转移。趋化因子对肿瘤的作用取决于多种因素，包括特定的趋化因子类型、肿瘤类型和微环境的特性。这些因子可能对肿瘤有促进作用，也可能有抑制作用。本研究中涉及到4种趋化因子，分别为CXCL3、CXCL5、CXCL8、CXCL13。CXCL3 主要由肿瘤细胞产生，对免疫细胞，特别是髓系源性抑制细胞（myeloid derived suppressor cells，MDSCs）有强烈的化学趋性作用[29]。研究发现，MDSCs 的聚集与多形性胶质母细胞瘤患

者的预后密切相关。这项研究还发现，CXCL3 的表达与多形性胶质母细胞瘤患者的预后密切相关，可以作为多形性胶质母细胞瘤有价值的预后标记物[30]。CXCL5 的过度表达增强了结直肠癌细胞的迁移和侵袭能力，这可能是通过激活特定途路引发上皮间质转化（epithelial-mesenchymal transition，EMT）实现的[31]。CXCL8 是主要的人类中性粒细胞趋化因子和炎症的主要介质。有研究证明 CXCL8 促进巨噬细胞 M2 极化，增加结直肠癌微环境中的免疫浸润[32]，肝癌患者血液中的 CXCL8 水平与肝癌微血管浸润密切相关[33]。CXCL13 蛋白及其受体 CXCR5 在炎症反应、免疫反应以及淋巴结新生中发挥了重要作用[34]，目前研究显示，肿瘤环境中的 CXCL13 对癌细胞的增殖、迁移和侵袭产生影响[35]。这些基因均涉及肿瘤的免疫微环境，且构建的神经网络模型对结直肠癌的发生预测具有较高的准确性，这对临床结直肠癌的诊疗具有重要意义，可以在患者炎症的早期阶段对患者进行指标监测，对患者炎癌转化的风险进行评估，以期对患者的可能癌变进行早期的干预。在后续炎癌转化研究中可以重点关注这些基因在结直肠癌发生发展中的作用。

TIMP1、*CXCL13* 和 *CXCL8* 3 个基因展现出和总生存期的相关性，*TIMP1* 高表达对患者是风险因素，*CXCL13* 和 *CXCL8* 高表达对患者来说是保护因素，后续可以重点研究这 3 个基因在炎癌转化过程中的主要机制。虽然机器学习和神经网络模型对生物靶点的预测得到了充分应用，但仍具有不足之处，二者对数据输入的质量具有严格要求，低质量的数据往往难以得到可靠的结果；神经网络模型的“黑箱”特性意味着其预测结果往往难以解释。

3.3 靶向预测的中药是干预结直肠癌炎癌转化过程的潜在药物

炎症在肿瘤发生发展过程中的作用目前已经得到了广泛的证实，其被列入“十大肿瘤特征”之一，有研究证实炎症在早期黏膜病变、结肠腺瘤、结肠癌过程中长期存在；在这个背景下，早期干预显得尤为重要，如果能在疾病尚处于发展初期阶段即炎症阶段，就进行积极的干预措施，那么结直肠癌的防治效果将可能显著提高。中医一直秉承“治未病”的理念，尤其强调在疾病尚处于发展的初期阶段实施积极的干预措施[36]，这一理念与本研究的目标——在结直肠癌的炎症阶段进行干预是一致的。

本研究对 7 个验证后的核心基因进行了中药的预测，对预测的中药分析后发现，寒性和温性中药所占的比例最大；苦味、辛味、甘味中药所占比例最大，表明这些中药对关键基因的影响可能更为显著。在功能主治的分类中，清热药和补虚药所占的比例最大，这与结直肠癌患者的中医证型不谋而合，1 项研究显示结直肠癌的中医证型以脾虚和湿热为主[37-38]。这可能暗示着这两类中药在调节关键基因及可能的生存影响中具有重要作用。本研究预测的清热药的代表药物有关木通、七叶一枝花、鱼腥草、黄芩、大黄等，已有研究证实部分中药在炎癌转化中发挥作用，关木通中的木通皂苷 D 通过激活核因子红细胞 2 相关因子 2 通路抑制炎症反应[39]，七叶一枝花提取物对多种肿瘤中具有抑制作用，包括肺癌[40]、乳腺癌[41]等；鱼腥草可以抑制促癌因子并抑制炎症反应，是一种典型的抗炎中药[42]；大黄在临床的应用十分广泛，对胰腺炎、胆囊炎等疾病具有显著的抗炎作用[43-44]。本研究预测的补虚药的代表药物有鹿骨、人参、淫羊藿、沙苑子等，人参是补气、补脾之品，现有研究亦显示人参能够调节炎症诱导的血管形成[45]，增强免疫功能。其他预测中药在炎癌转化中的机制研究较少，但可能亦对炎癌转化过程具有干预作用。

中国传统医学的治未病理论强调预防疾病的发生，这在结直肠癌的炎癌转化过程中具有极其重要的指导意义。对于结直肠癌而言，炎症状态可能是癌症发生的早期阶段，这与本研究中发现的炎癌转化过程中免疫浸润的重要作用相符。这为今后的研究提供了一个理论基础，即在免疫浸润的过程中寻找早期干预的机会，从而阻止或延缓结直肠癌的发展。通过机器学习算法和神经网络模型的应用，可以更深入地了解结直肠癌的炎癌转化过程，并找到更有效的预防和治疗方法；这不仅可以提高结直肠癌的早期诊断率，还能为结直肠癌患者提供更个性化的治疗方案。

利益冲突 所有作者均声明不存在利益冲突

参考文献

[1] 肖祥, 吴宣諭. 胰腺癌“炎-癌”转化关键基因筛选及干预中药的预测分析 [J]. 中草药, 2022, 53(15): 4795-4806.

[2] 肠外肠内营养学分会指南与规范编委会. 神经系统疾病与营养支持 [J]. 中国临床营养杂志, 2008, 16(5): 265-267.

[3] Hanahan D. Hallmarks of cancer: New dimensions [J]. *Cancer Discov*, 2022, 12(1): 31-46.

[4] 金光亮. 《内经》未病概念与“治未病”理论探讨 [J]. 北京中医药大学学报, 2006, 29(12): 804-806.

[5] Leek J T, Johnson W E, Parker H S, *et al*. The sva package for removing batch effects and other unwanted variation in high-throughput experiments [J]. *Bioinformatics*, 2012, 28(6): 882-883.

[6] Ritchie M E, Phipson B, Wu D, *et al*. Limma powers differential expression analyses for RNA-sequencing and microarray studies [J]. *Nucleic Acids Res*, 2015, 43(7): e47.

[7] Love M I, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2 [J]. *Genome Biol*, 2014, 15(12): 550.

[8] Gao C H, Yu G C, Cai P. ggVennDiagram: An intuitive, easy-to-use, and highly customizable R package to generate Venn diagram [J]. *Front Genet*, 2021, 12: 706907.

[9] Wickham H. ggplot2 [J]. *WIREs Comp Stat*, 2011, 3(2): 180-185.

[10] Yu G C, Wang L G, Han Y Y, *et al*. ClusterProfiler: An R package for comparing biological themes among gene clusters [J]. *OMICS*, 2012, 16(5): 284-287.

[11] Chin C H, Chen S H, Wu H H, *et al*. CytoHubba: Identifying hub objects and sub-networks from complex interactome [J]. *BMC Syst Biol*, 2014, 8(Suppl 4): S11.

[12] 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.

[13] Schmitt M, Greten F R. The inflammatory pathogenesis of colorectal cancer [J]. *Nat Rev Immunol*, 2021, 21(10): 653-667.

[14] Jiang X J, Wang J, Deng X Y, *et al*. Role of the tumor microenvironment in PD-L1/PD-1-mediated tumor immune escape [J]. *Mol Cancer*, 2019, 18(1): 10.

[15] Kriegeskorte N, Golan T. Neural network models and deep learning [J]. *Curr Biol*, 2019, 29(7): R231-R236.

[16] Crocker M, Ashley S, Giddings I, *et al*. Serum angiogenic profile of patients with glioblastoma identifies distinct tumor subtypes and shows that TIMP-1 is a prognostic factor [J]. *Neuro Oncol*, 2011, 13(1): 99-108.

[17] Huang R, Wang K J, Gao L, *et al*. *TIMP1* is a potential key gene associated with the pathogenesis and prognosis of ulcerative colitis-associated colorectal cancer [J]. *Onco Targets Ther*, 2019, 12: 8895-8904.

[18] Dittmer A, Dittmer J. A CAF-fueled TIMP-1/CD63/ITGB1/STAT3 feedback loop promotes migration and growth of breast cancer cells [J]. *Cancers*, 2022, 14(20): 4983.

[19] Kumar J D, Aolymat I, Tiszlavicz L, *et al*. Chemerin acts via CMKLR1 and GPR1 to stimulate migration and invasion of gastric cancer cells: Putative role of decreased TIMP-1 and TIMP-2 [J]. *Oncotarget*, 2019, 10(2): 98-112.

[20] Schoeps B, Eckfeld C, Prokopchuk O, *et al*. TIMP1 triggers neutrophil extracellular trap formation in pancreatic cancer [J]. *Cancer Res*, 2021, 81(13): 3568-3579.

[21] Liu L, Chen Z W, Yi S. Association between the expression of *MMP1* gene and prognosis in head and neck squamous cell carcinoma [J]. *J Clin Otorhinolaryngol Head Neck Surgery*, 2018, 32(4): 287-291.

[22] Wen J, Wang H, Yang B, *et al*. Expression of matrix metalloproteinase 1 in liver cancer and its relationship with migration and invasion [J]. *Chin J Experim Surgery*, 2020, 37(3): 534-536.

[23] Peng Z M, Yang Z L, Liu Y W. Expression of MMP1 and TIMP1 proteins in lung cancer and its biological significance [J]. *Bull Hunan Med Univ*, 2002, 27(2): 159-161.

[24] Mao S, Xia A L, Tao X W, *et al*. A pan-cancer analysis of the prognostic and immunological roles of matrix metalloprotease-1 (MMP1) in human tumors [J]. *Front Oncol*, 2022, 12: 1089550.

[25] Yu J H, Xu Z S, Guo J, *et al*. Tumor-associated macrophages (TAMs) depend on MMP1 for their cancer-promoting role [J]. *Cell Death Discov*, 2021, 7(1): 343.

[26] Lièvre A, Milet J, Carayol J, *et al*. Genetic polymorphisms of MMP1, MMP3 and MMP7 gene promoter and risk of colorectal adenoma [J]. *BMC Cancer*, 2006, 6: 270.

[27] Alim M A, Njenda D, Lundmark A, *et al*. Pleckstrin levels are increased in patients with chronic periodontitis and regulated via the MAP kinase-p38α signaling pathway in gingival fibroblasts [J]. *Front Immunol*, 2021, 12: 801096.

[28] Oboshi W, Watanabe T, Hayashi K, *et al*. QPY/RAH haplotypes of the *GZMB* gene are associated with natural killer cell cytotoxicity [J]. *Immunogenetics*, 2018, 70(1): 29-36.

[29] Russo E, Santoni A, Bernardini G. Tumor inhibition or tumor promotion? The duplicity of CXCR3 in cancer [J]. *J Leukoc Biol*, 2020, 108(2): 673-685.

[30] Codrici E, Popescu I D, Tanase C, *et al*. Friends with benefits: Chemokines, glioblastoma-associated microglia/macrophages, and tumor microenvironment [J]. *Int J Mol Sci*, 2022, 23(5): 2509.

[31] Zhang S N, Ding J L, Wang J Y, *et al*. CXCL5 downregulation in villous tissue is correlated with recurrent spontaneous abortion [J]. *Front Immunol*, 2021, 12: 717483.

[32] Gao S, Chai X, Liu H, et al. CXCL8 promotes M2 macrophage chemotaxis and infiltration in colorectal cancer microenvironment [J]. *Chin J Biochem Mole Biol*, 2022, 38(4): 495-504.

[33] Liu H, Bai Y, Li F L, *et al*. Combined serum CXCL8, CXCL9 and CXCL13 tests for the prediction of microvascular invasion in hepatocellular carcinoma [J]. *Biomark Med*, 2023, 17(5): 265-272.
[34] Hsieh C H, Jian C Z, Lin L I, *et al*. Potential role of CXCL13/CXCR5 signaling in immune checkpoint inhibitor treatment in cancer [J]. *Cancers*, 2022, 14(2): 294.
[35] Hussain M, Liu J, Wang G Z, *et al*. CXCL13 signaling in the tumor microenvironment [J]. *Adv Exp Med Biol*, 2021, 1302: 71-90.
[36] 索凤茹. 治未病与肿瘤防治 [J]. 实用中医内科杂志, 2014, 28(5): 24-25.
[37] 林胜友, 沈敏鹤, 舒静娜, 等. 780 例大肠癌中医证型分类特征的回顾性分析 [J]. 浙江中医药大学学报, 2011, 35(3): 322-324.
[38] 邱艳艳, 梁波, 胡送娇, 等. 大肠癌中医证型分布特点的现代文献研究 [J]. 上海中医药大学学报, 2014, 28(6): 52-55.
[39] Luo J F, Zhou H, Lio C K. *Akebia* saponin D inhibits the inflammatory reaction by inhibiting the IL-6-STAT3-DNMT3b axis and activating the Nrf2 pathway [J]. *Molecules*, 2022, 27(19): 6236.
[40] He H, Zheng L, Sun Y P, *et al*. Steroidal saponins from *Paris polyphylla* suppress adhesion, migration and invasion of human lung cancer A549 cells via down-regulating MMP-2 and MMP-9 [J]. *Asian Pac J Cancer Prev*, 2014, 15(24): 10911-10916.
[41] Lee M S, Yuet-Wa J C, Kong S K, *et al*. Effects of polyphyllin D, a steroidal saponin in *Paris polyphylla*, in growth inhibition of human breast cancer cells and in xenograft [J]. *Cancer Biol Ther*, 2005, 4(11): 1248-1254.
[42] Wang J F, Dempsey E, Corr S C, *et al*. The traditional Chinese medicine *Houttuynia cordata* Thunb. Decoction alters intestinal barrier function via an EGFR dependent MAPK (ERK1/2) signalling pathway [J]. *Phytomedicine*, 2022, 105: 154353.
[43] 曹强, 郭亚菲, 叶蕾蕾, 等. 大黄及其活性成分抗炎作用及机制的研究进展 [J]. 中草药, 2022, 53(22): 7285-7294.
[44] 胡蓝烁, 黄炫淳, 杨先照. 大黄相关复方联合西医常规治疗急性胆囊炎的 Meta 分析 [J]. 中国中西医结合外科杂志, 2022, 28(6): 830-838.
[45] Dai D, Zhang C F, Williams S, *et al*. Ginseng on cancer: Potential role in modulating inflammation-mediated angiogenesis [J]. *Am J Chin Med*, 2017, 45(1): 13-22.

[责任编辑　潘明佳]