

# 结直肠“正常黏膜-腺瘤-腺癌”序列转化的靶向基因筛选及防治中药预测

季双双<sup>1,2</sup>, 龙思丹<sup>2</sup>, 杨洋<sup>1,2</sup>, 张书信<sup>1\*</sup>, 田明<sup>3\*</sup>

1. 北京中医药大学东直门医院, 北京 100700

2. 北京中医药大学研究生院, 北京 100029

3. 中国中医科学院西苑医院, 北京 100091

**摘要:** **目的** 探究结直肠腺癌经典序列转化中的差异基因及其与免疫细胞浸润和患者生存时间的相关性, 并预测具有靶向作用的中药。**方法** 从 GEO 数据库寻找符合要求的配对样本数据, 使用 R 语言筛选序列转化过程中始终存在差异表达的关键基因, 并通过基因本体 (gene ontology, GO) 进行功能注释。利用基因表达谱互动分析数据库 (Gene Expression Profiling Interactive Analysis, GEPIA) 和人类蛋白质图谱数据库验证关键基因在转录组和蛋白组水平的表达。通过 Kaplan-Meier Plotter 和 TIMER 数据库分别进行基因表达水平与结直肠癌患者生存时间、免疫细胞浸润的相关性分析。最后, 利用 Coremine Medical、BATMAN-TCM 和草本植物数据库进行中药预测, 使用 Cytoscape 构建“基因-中药”靶向网络, 综合 Network Analyzer 和 Swiss ADME 数据库评估活性成分分类药性, PyMOL 软件进行分子对接。**结果** 获得结直肠腺癌经典序列转化的关键差异基因 48 个, 筛选得到 9 个核心基因: TAGLN、FN1、ACTA2、DCN、REG4、GUCA2A、SPINK4、DEFA6、ZG16。与正常组织相比, 结直肠癌组织中的核心基因在转录组和蛋白质水平上均具有统计学差异。功能性分析提示其主要参与肠道免疫、细胞外基质构成等过程。相关性分析表明核心基因的表达水平能够影响结直肠癌患者的总生存期和无复发生存期, CD4<sup>+</sup>T 细胞、巨噬细胞、中性粒细胞和树突细胞是被核心基因调节的主要免疫细胞群。获得靶向中药 129 味, 包括 36 味药食同源中药和 27 味毒性中药, 以补益、清热、化痰为主, 性味以寒、平、温、甘、苦、辛居多, 多归肝、肺、脾经, 其中葡萄、金荞麦、半夏、地龙、肉苁蓉的中心度值较高, 筛选得到 6 种关键活性成分缬氨酸、鼠李糖、脯氨酸、橙花醛、蛋氨酸、甘氨酸, 分子对接显示中药小分子活性成分与靶向蛋白质对接稳定。**结论** 筛选出“正常黏膜-肠腺瘤-肠腺癌”序列转化中始终差异表达的关键疾病基因, 这些基因可能通过参与免疫调节、细胞外基质等过程影响结直肠癌患者的远期预后, 金荞麦、葡萄、肉苁蓉等药食同源中药有望成为新药开发来源, 为预防肠道早癌和癌症进展提供新的思路。

**关键词:** 结直肠癌; 肠腺瘤; 生物信息学; 分子对接; 金荞麦; 半夏; 肉苁蓉; 缬氨酸; 鼠李糖; 脯氨酸; 橙花醛; 蛋氨酸; 甘氨酸

中图分类号: Q811.4; R285 文献标志码: A 文章编号: 0253-2670(2025)03-0933-14

DOI: 10.7501/j.issn.0253-2670.2025.03.019

## Identification of targeted genes in process of colorectal “mucosa-adenoma-adenocarcinoma” sequence and prediction of traditional Chinese medicine for prevention and treatment

Ji Shuangshuang<sup>1,2</sup>, LONG Sidan<sup>2</sup>, YANG Yang<sup>1,2</sup>, ZHANG Shuxin<sup>1</sup>, TIAN Ming<sup>3</sup>

1. Dongzhimen Hospital, Beijing University of Chinese Medicine, Beijing 100700, China

2. Graduate School, Beijing University of Chinese Medicine, Beijing 100029, China

3. Xiyuan Hospital, China Academy of Chinese Medical Science, Beijing 100091, China

**Abstract: Objective** To identify differentially expressed genes (DEGs) in the process of colorectal mucosa-adenoma-adenocarcinoma sequence and their relationships with immune cell infiltration and survival time of patients with colorectal cancer (CRC), and to predict potential targeted traditional Chinese medicines (TCMs). **Methods** The paired data was downloaded from

收稿日期: 2024-07-12

基金项目: 北京市自然科学基金项目 (7242237)

作者简介: 季双双, 博士研究生, 研究方向为中医药防治大肠肛门疾病。E-mail: 522876012@qq.com

\*通信作者: 张书信, 男, 博士, 主任医师、博士生导师, 研究方向肛肠外科。E-mail: 13661027611@126.com

田明, 副教授, 副主任医师, 研究方向胃肠外科。E-mail: tianming\_33@sina.com

the GEO database to obtain DEGs using R. Biological pathways were explored through gene ontology (GO). Expression levels of DEGs were validated using GEPIA and the Human Protein Atlas. Kaplan-Meier Plotter and TIMER databases were used to assess the correlation between DEGs and survival time, as well as immune cell infiltration. Finally, Coremine Medical, BATMAN, and Herb databases were used to predict TCMs. A “Gene-TCM” network was constructed with Cytoscape. Combined with Network Analyzer, the drug-likeness of the active compounds was evaluated using Swiss ADME databases. Molecular docking was performed with PyMOL software. **Results** Forty-eight key DGEs were identified in transformation, with nine hub genes: TAGLN, FN1, ACTA2, DCN, REG4, GUCA2A, SPINK4, DEFA6, and ZG16. At both at the transcriptome and proteome levels, the hub genes exhibited statistically significant differences in expression between tumor and normal tissues. Functional analysis indicated they were involved in intestinal immunity and extracellular matrix composition. Correlation analysis suggested that the expression levels of hub genes can affect the overall survival (OS) and relapse-free survival (RFS) of patients with CRC. CD4<sup>+</sup> T cells, macrophages, neutrophils, and dendritic cells were primary immune cells regulated by hub genes. There were 130 potential TCMs, including 36 Medicine-food homologous TCMs and 27 toxic TCMs. The efficacy was mainly to tonify, clear heat, and resolve phlegm. Moreover, their four natures and five flavors were primarily cold, neutral, warm, sweet, bitter, and pungent. Most of them belonged to the liver, lung, and spleen meridians. Putao (*Vitis vinifera*), Jinqiaomai (*Fagopyri Dibotryis Rhizoma*), Banxia (*Pinelliae Rhizoma*), Dilong (*Pheretima*), Roucongrong (*Cistanches Herba*) had the strongest connections to hub genes. Six key active compounds were valine, L-rhamnose, proline, neral, D-methionine, and glycine. Molecular docking showed that they were stable in docking with markers. **Conclusion** This study identified genes that were consistently differentially expressed during the sequence. These DEGs may affect the prognosis of patients with CRC by regulating immune cells and extracellular matrix. Medicine-food homologous TCMs such as *Fagopyri Dibotryis Rhizoma*, *Vitis vinifera*, and *Cistanches Herba* could be promising sources for drug development. This study sheds new light on the prevention of colorectal adenoma and cancer.

**Key words:** colorectal cancer; colorectal adenoma; bioinformatics; molecular docking; *Fagopyri Dibotryis Rhizoma*; *Pinelliae Rhizoma*; *Cistanches Herba*; valine; L-rhamnose; proline; neral; D-methionine; methionine; glycine

结直肠癌 (colorectal cancer) 是发病率和死亡率排名第 2 的恶性肿瘤, 60%~65% 是非遗传的散发性结直肠 (sporadic colorectal cancer, sCRC), 全球负担严重<sup>[1]</sup>。绝大部分 sCRC (85%~90%) 是通过“正常黏膜-腺瘤-腺癌”序列进展而来<sup>[2]</sup>, 该序列被认为是 sCRC 发展的经典序列, 整个过程平均需要 10~15 年<sup>[3]</sup>, 为早期诊断和早期防治提供了可能。结直肠镜检查在该阶段起到了关键性作用, 但仍存在受检率低下<sup>[4]</sup>、漏诊<sup>[5]</sup>、并发症<sup>[6]</sup>、成本效应<sup>[7]</sup>等问题, 并且该治疗不能改变肿瘤微环境, 术后复发率<sup>[8]</sup>无显著性改善。在预防药物方面, 阿司匹林等非甾体类药物的观察时间较短、证据等级不足, 并且存在副作用, 因此尚未推广<sup>[9]</sup>。

中药以多靶点的天然优势能通过改善复杂的微环境预防肿瘤的进展和转移, 值得深入挖掘。目前已有多项临床试验<sup>[10]</sup>、动物实验<sup>[11]</sup>、文献综述<sup>[12]</sup>研究中药及复方预防肠道肿瘤的疗效和机制, 其组方用药依据多来源于中医药古籍记载和临床经验, 因此, 其疗效和机制可能受到多种因素的影响。随着新一代测序技术的出现, 基因表达谱分析成为了解疾病发生和进展的有效方法。目前, 已开展了基于肠癌<sup>[13]</sup>和炎癌转化<sup>[14]</sup>中差异

基因的筛选和中药预测研究, 少数研究也关注到了序列转化中的基因改变, 但边杰等<sup>[15]</sup>只关注了肠癌中显著表达而在腺瘤中不显著表达的差异基因, Pan 等<sup>[16]</sup>只关注了转录因子的改变。因此, 针对结直肠癌经典序列转化的全程差异基因和中药挖掘工作一直尚未展开。为此, 本研究利用基因芯片数据库 (Gene Expression Omnibus, GEO), 选择同时患有肠腺瘤的结直肠腺癌患者, 利用其个体配对样本 (癌旁黏膜组织-腺瘤组织-腺癌组织) 进行生物信息学分析, 首次尝试在序列转化中挖掘出始终存在差异表达的疾病基因, 这些基因被认为与疾病进展和患者预后密切相关。通过预测靶向这些基因的中药能够为结直肠肿瘤的全程防治提供用药选择和依据, 并利用分子对接技术进行验证, 相关蛋白质和化合物的结构对接有可能为开发新药提供理论基础。

## 1 材料与方法

### 1.1 关键基因筛选

以“colorectal adenocarcinoma”为关键词检索 GEO 基因表达综合数据库, 数据集需同时满足: (1) 阵列表达谱分析; (2) 智人 (*Homo sapiens*); (3) 肠腺瘤、肠腺癌、癌旁黏膜 (正常肠黏膜) 组织均

来自同一患者的成对匹配样本；(4)全基因表达谱；(5)非遗传性。最终筛选得到 GSE117606 基因芯片为本研究分析的对象。

数据集特征：共 70 位患者的 208 个肠组织样本，包括 69 个腺瘤、65 个癌旁组织和 74 个腺癌，其中癌组织样本包含 18 个 I 期、24 个 II 期、26 个 III 期、3 个 IV 期和 3 个无分期数据的样本（TNM 分期标准）。样本使用 Affymetrix HT HG-U133+PM Array 平台进行分析。

首先，利用 GEO2R 工具<sup>[17]</sup>筛选得到 2 组差异表达基因：(1) 癌旁 vs 肠腺瘤；(2) 肠腺瘤 vs 肠腺癌，筛选条件为  $|\log_2FC| \geq 1$ 、校正  $P < 0.05$ 。然后，利用 R 语言的“Venn”插件包筛选上述 2 组间差异基因的交集基因，该交集基因被认为是全程参与“肠黏膜-腺瘤-癌”序列的关键基因。最后，使用 GeneCards 数据库<sup>[18]</sup> (<https://www.genecards.org/>) 获取基因符号、功能等基因基本信息，利用 R 语言的“Heatmap”插件包绘制热图，将关键基因在 3 种组织样本中的表达情况可视化。

## 1.2 核心基因筛选及功能注释

将上述得到的关键基因导入 STRING 数据库<sup>[19]</sup> (<https://www.genecards.org/>) 以进一步筛选重要的核心基因，筛选条件为中等置信度 0.4，获取蛋白互作（protein-protein interaction, PPI）网络模型。然后，将结果导入 Cytoscape 软件<sup>[20]</sup>，使用“CytoNCA”插件中的 degree（without weight）方法分析节点中心性，“cyto Hubba”插件中的 MCC 方法筛选核心基因，并进行可视化。核心基因的功能注释使用 Metascape 数据库<sup>[21]</sup> (<https://metascape.org/>) 进行基因本体论（gene ontology, GO）分析，并使用 R 软件完成数据可视化。

## 1.3 核心基因表达水平验证

将上述得到的核心基因在不同的数据集中考察表达水平以验证上述生物信息分析的结果。在基因表达谱互动分析数据库（Gene Expression Profiling Interactive Analysis, GEPIA）<sup>[22]</sup> (<http://gepia.cancer-pku.cn/>) 中查看核心基因分别在正常肠组织、结肠腺癌（colon adenocarcinoma, COAD）、直肠腺癌（rectum adenocarcinoma, READ）中的转录组表达水平，其肿瘤组织和正常组织样本分别来自 TCGA 数据库 (<https://www.genome.gov/Funded-Programs-Projects/Cancer-Genome-Atlas>)<sup>[23]</sup> 和 GTEx 数据库 (<https://gtexportal.org/home/>)<sup>[24]</sup>。

人类蛋白质图谱数据库（<https://www.proteinatlas.org/>）<sup>[25]</sup> 以免疫组织化学染色的组织样本图像的形式显示蛋白质表达。在该数据库中查看核心基因在正常肠组织、COAD 和 READ 组织样本中的蛋白质组表达水平。

## 1.4 核心基因表达与生存曲线、肿瘤免疫细胞浸润的相关性

使用 Kaplan-Meier Plotter 数据库<sup>[26]</sup> (<http://kmplot.com/analysis/>) 中的 Kaplan-Meier 曲线查看核心基因的表达对 COAD 和 READ 患者的总生存率（overall survival, OS）和无复发生存率（relapse-free survival, RFS）的影响。Kaplan-Meier 绘图仪是一款功能强大的在线工具，可用于评估 54 000 个基因对 21 种不同癌症类型生存的影响。计算了对数秩  $P$  值（Logrank  $P$ ）和风险比（hazard ratio, HR）以及 95% 置信区间（confidence interval, CI），Logrank  $P < 0.05$  具有统计学意义。

利用 TIMER 数据库<sup>[27]</sup> (<https://cistrome.shinyapps.io/timer/>) 查看基因与 6 种主要肿瘤相关免疫细胞之间的相关性，包括 B 细胞、CD4<sup>+</sup>T 细胞、CD8<sup>+</sup>T 细胞、树突状细胞、巨噬细胞、中性粒细胞。TIMER 数据库包含了来自 32 种癌症类型的 10 897 个样本，用于评估免疫浸润的数量，使其成为系统研究不同癌症类型免疫浸润的合适资源。进一步确定了核心基因表达水平与肿瘤纯度之间的关系，以减少分析偏差。

## 1.5 中药预测

将核心靶基因分别输入 Coremine Medical (<https://www.coremine.com/medical/>)、BATMAN-TCM (<http://bionet.ncpsb.org.cn/batman-tcm/>) 和草本植物数据库 (<http://herb.ac.cn/>)，这些证据来源于不同数据库中的小鼠和智人的高通量测序和微阵列实验的原始数据<sup>[28]</sup>。去除各种草药中具有广泛作用且常见的成分，设置筛选条件为  $P < 0.01$ ，即得到具有靶向作用的中药。综合《中国药典》《中药大辞典》《神农本草经》等资料，考察和规范预测得到的中药名称及四气、五味、归经、功效等属性。国家卫生健康委员会 2018 年发布的最新文件《国卫食品便民函（2018）8 号》中指明 110 种中药被认定为药品和食品成分，并被推广用于适当医疗用途<sup>[29]</sup>，基于该文件考察预测中药是否属于药食同源类，利用 Excel 对中药特征进行归纳整理和可视化。利用 Cytoscape<sup>[20]</sup> 制作“基因-中药”靶向网络，筛选度值

排名前 5 的中药作为核心中药进行后续分析。

### 1.6 活性成分筛选及分子对接

将上述得到的核心中药导入本草组鉴草本植物数据库 (<http://herb.ac.cn/>), 查询化学成分。(1) 初筛活性成分: 将核心中药的所有化学成分通过 Network Analyzer 推算拓扑参数, 参考中心度 (betweenness centrality)、紧密中心度 (closeness centrality), 以节点连接度值 (degree) 进行初筛, 得到排序前 15 位的核心活性成分。(2) 复筛活性成分: 参考中医药系统药理数据库和分析平台 (TCMSP) [30] 的口服生物利用度和类药性 [31], 将成分导入 Swiss ADME 数据库 (<http://www.swissadme.ch>) 进行类药性评估, 以肠道通透性为 “High”, 类药性五原则至少有 3 个 “Yes” 进行复筛确定关键成分。

从 PubChem 数据库 [32] (<https://pubchem.ncbi.nlm.nih.gov/>) 中下载药物活性成分化合物的二维结构, 并将其保存为 “SDF” 格式。使用 Open Babel 2.3.2 软件将 SDF 文件转化为 PDB 文件。从 The Protein Data Bank (PDB) 数据库 [33] 下载核心模块的受体蛋白质三维结构, 并以 “PDB” 格式保存为蛋白质受体。使用 PyMOL 软件 (2.3.6 版) 去除水分子, 从核心目标蛋白中分离出原始配体。处理后的

蛋白质目标被导入 AutoDock 软件 (4.2.0 版) [34] 进行氢化处理, 平衡电荷等修饰。配体小分子和受体蛋白转为 pdbqt 格式保存。用 AutoDock-Vina 软件 (1.1.2 版) [35] 进行分子对接, 对受体-配体复合物进行综合评分, 结合能最低、聚簇频率最高的对接构象被认为是配体与蛋白质之间最潜在的结合模式。分子对接可视化使用 PyMOL 软件。

## 2 结果

### 2.1 关键差异基因筛选

通过比较正常肠黏膜和肠腺瘤数据集, 共筛选获得差异基因 281 个; 通过比较肠腺瘤和肠腺癌数据集, 共筛选获得差异基因 191 个。Venn 图显示以上差异基因中存在 48 个重复的基因 (图 1-A), 这些基因在 “正常黏膜-腺瘤-癌” 序列演变中全程参与, 并且其表达量始终保持着 2 倍以上的显著差异, 被认为是全程参与序列转化和疾病进展的关键差异基因。48 个关键差异基因在 3 种组织样本中的表达水平分布见图 1-B。

### 2.2 核心差异基因筛选及功能注释

将分析结果文件导入 Cytoscape 软件进行可视化, 图 2 显示共有 33 个节点, 128 条边。经 Cyto Hubba 计算列取前 9 位作为核心基因, 按重要程度依次排序为转运蛋白 (transgelin, TAGLN)、纤连

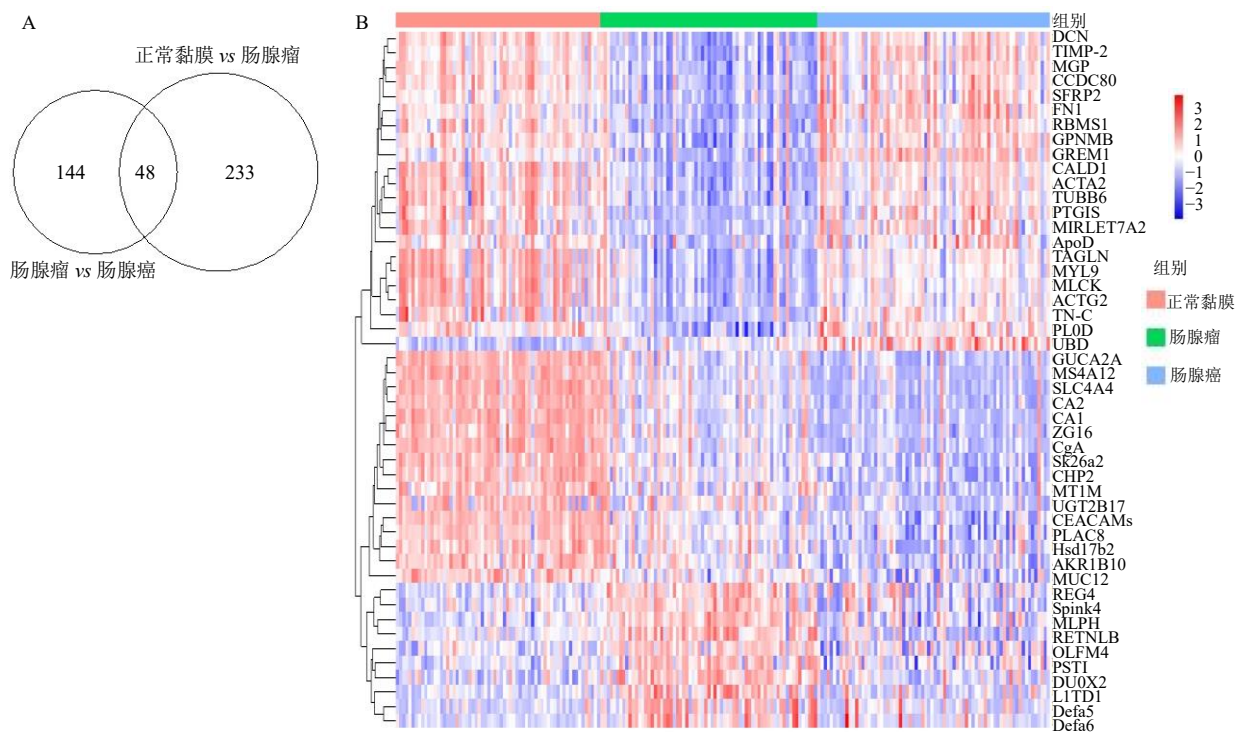
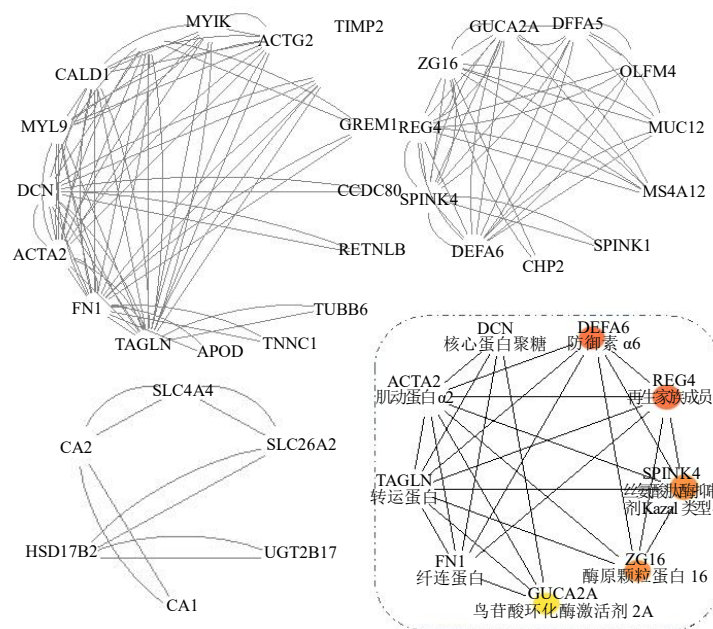


图 1 关键差异基因筛选 Venn 图 (A) 及表达水平分布 heatmap (B)

Fig. 1 Venn map (A) and heat map (B) of key differentially expressed genes



颜色越深代表该节点的度值越大，重要程度越高；虚线方框内为筛选的核心差异基因。

Darker the color the greater the degree value of the node, and the higher the importance. The core genes shown in the dotted box.

图 2 PPI 网络模型

Fig. 2 PPI network model

蛋白 (fibronectin 1, FN1)、肌动蛋白  $\alpha 2$  (actin alpha 2, ACTA2)、核心蛋白聚糖 (decorin, DCN)、再生家族成员 4 (regenerating islet-derived family member 4, REG4)、鸟苷酸环化酶激活剂 2A (guanylate cyclase activator 2A, GUCA2A)、丝氨酸肽酶抑制剂 Kazal 类型 4 (serine peptidase inhibitor Kazal type 4, SPINK4)、防御素  $\alpha 6$  (defensin alpha 6, DEFA6)、酶原颗粒蛋白 16 (zymogen granule protein 16, ZG16)。

GO 分析包含生物过程 (biological process, BP)、细胞组分 (cellular component, CC)、分子功能 (molecular function, MF) 3 大方面，结果如图 3 所示。其中，BP 分析结果显示核心基因主要参与对细菌存在而引发的反应；CC 分析结果显示核心基因参与组成高尔基体腔和细胞外基质 (含胶原)；MF 分析结果显示核心基因与糖胺聚糖结合、肽酶调节活性、碳水化合物结合、酶激活等 MF 有关。

### 2.3 核心差异基因表达验证

为了验证筛选出来的核心基因是否具有代表性和治疗潜力，首先在 GEPIA 数据库中查看了 9 个基因的转录组表达，验证数据集包括 COAD 样本 275 例，匹配的正常样本 349 例，READ 样本 92 例，匹配正常样本 318 例。结果显示，与正常肠组织相

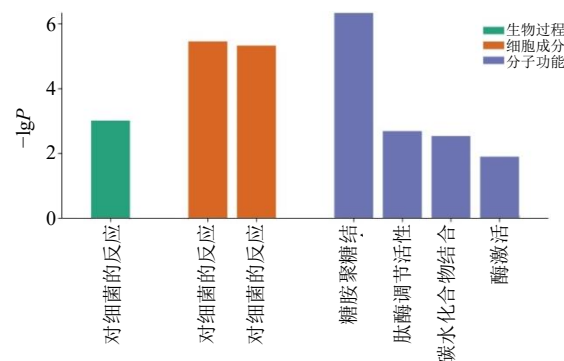


图 3 GO 富集分析

Fig. 3 GO enrichment analysis

比，核心基因 *TAGLN*、*FN1*、*ACTA2*、*DCN*、*ZG16* 在 COAD 和 READ 组织中的表达水平均显著降低 (图 4-A~D、I)，*GUCA2A* 基因表达在 COAD 组织中显著降低，但在 READ 组织中无明显改变 (图 4-F)。其他基因 *REG4*、*SPINK4*、*DEFA6* 在 COAD 和 READ 肿瘤组织中的表达水平均显著增高 (图 4-E、G~H)。人类蛋白质图谱数据库的免疫组织化学结果显示，与正常肠组织样本相比，上述 9 大核心基因在蛋白质组的表达水平与转录组结果相似 (图 5)。因此，该部分结果一定程度上验证了此前生信分析的结果，并突出展示了关键差异基因在结直肠癌组织中的变化情况。



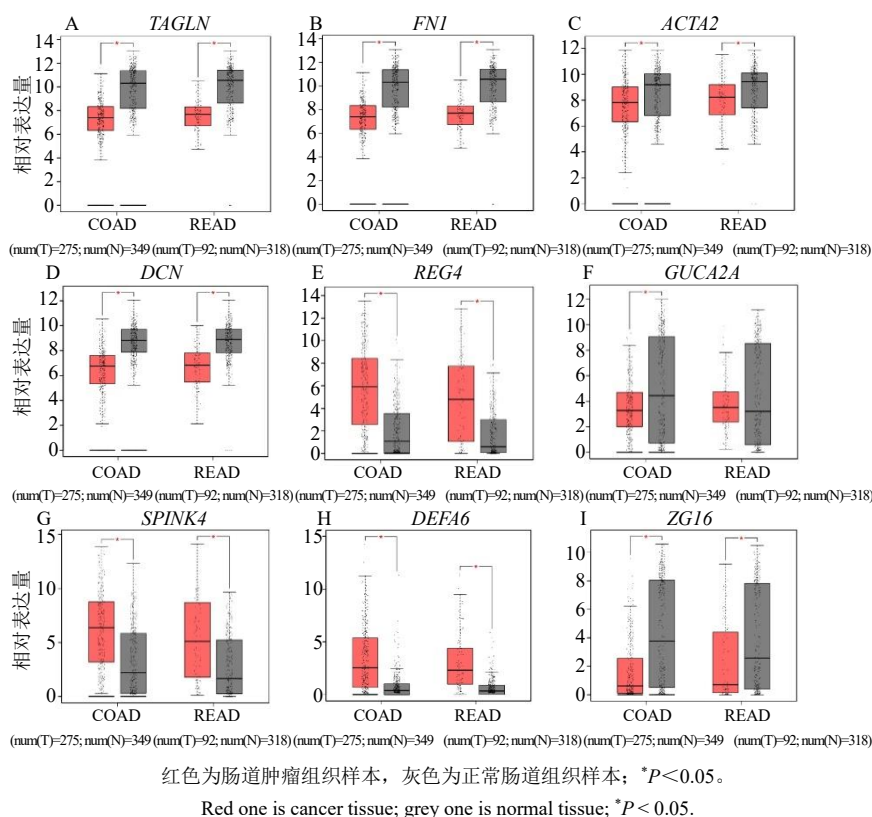


图 4 核心基因 (A—I) 的转录组表达水平  
Fig. 4 Transcriptome expression levels of core genes (A—I)

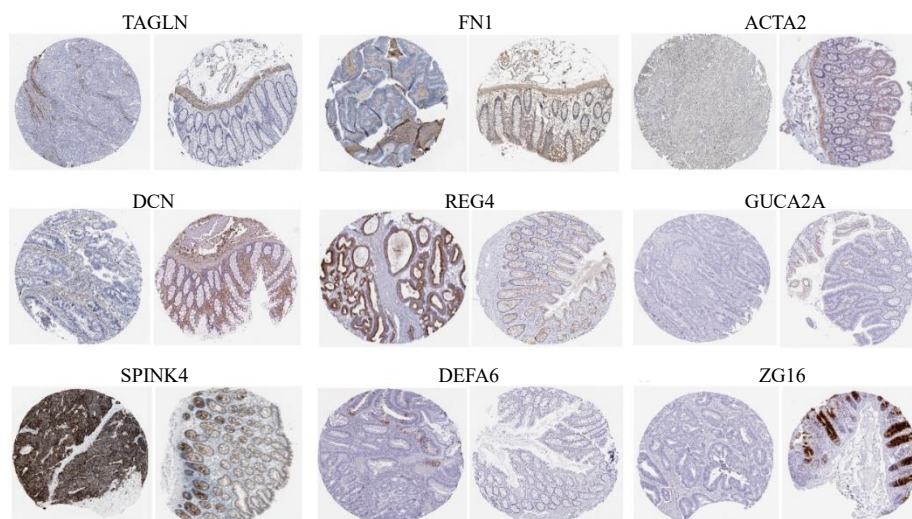


图 5 核心基因的蛋白表达水平  
Fig. 5 Protein expression levels of core genes

## 2.4 核心差异基因与生存时间的相关性

上述通过筛选和验证得到的 9 个核心差异基因被认为参与序列转化全程，与肠腺癌发病密切相关。通过考察这类基因表达水平与患者生存时间之

间的相关性，可以进一步揭示这类基因对肠腺癌疾病进展的作用。对结直肠癌患者的生存曲线数据进行提取，核心基因分别对 COAD 与 READ 患者的 OS 和 RFS 的影响见表 1。

表 1 核心基因和生存时间的相关性

Table 1 Correlation between core genes and survival time

| 基因符号   | 总生存期 (OS)        |                       |                  |                   | 无复发生存期 (RFS)     |                        |                           |                   |
|--------|------------------|-----------------------|------------------|-------------------|------------------|------------------------|---------------------------|-------------------|
|        | COAD<br>HR (95%) | COAD<br>Logrank P     | READ<br>HR (95%) | READ<br>Logrank P | COAD<br>HR (95%) | COAD<br>Logrank P      | READ<br>HR (95%)          | READ<br>Logrank P |
| TAGLN  | 1.48 (1.21~1.81) | $1 \times 10^{-4*}$   | 2.15 (0.86~5.36) | 0.094             | 2.62 (2.06~0.32) | $2.5 \times 10^{-16*}$ | 2.38 (0.27~0.68)          | 0.42              |
| FN1    | 1.45 (1.19~1.78) | 0.000 26*             | 0.68 (0.27~1.69) | 0.400             | 1.71 (1.38~2.11) | $5.6 \times 10^{-7*}$  | 307 307 223.25<br>(0~Inf) | 0.18              |
| ACTA2  | 1.58 (1.28~1.94) | $1.4 \times 10^{-5*}$ | 2.08 (0.83~5.19) | 0.110             | 1.97 (1.59~2.43) | $1.8 \times 10^{-10*}$ | 2.38 (0.48~11.89)         | 0.27              |
| DCN    | 1.33 (0.08~0.64) | 0.007 1*              | 0.50 (0.23~1.08) | 0.074             | 1.57 (1.20~2.05) | 0.000 96*              | 0.27 (0.04~1.96)          | 0.17              |
| REG4   | 0.87 (0.66~0.14) | 0.31                  | 0.05 (0.25~1.22) | 0.140             | 1.28 (1.02~1.61) | 0.033*                 | 0 (0~Inf)                 | 0.08              |
| GUCA2A | 0.79 (0.63~0.99) | 0.041*                | 0.54 (0.23~0.25) | 0.140             | 1.13 (0.92~1.40) | 0.250                  | 0.46 (0.09~2.33)          | 0.34              |
| SPINK4 | 0.80 (0.65~0.98) | 0.033*                | 0.39 (0.18~0.84) | 0.014*            | 0.82 (0.66~1.01) | 0.063                  | 0 (0~Inf)                 | 0.13              |
| DEFA6  | 0.86 (0.70~1.05) | 0.14                  | 0.44 (0.20~0.96) | 0.034*            | 0.79 (0.64~0.97) | 0.026*                 | 0.36 (0.07~1.97)          | 0.22              |
| ZG16   | 0.77 (0.63~0.94) | 0.011*                | 0.46 (0.21~1.00) | 0.045*            | 0.78 (0.63~0.96) | 0.020*                 | 0.14 (0.03~0.78)          | 0.009 2*          |

HR-危险比, HR=1 表示无效应; HR<1 表示风险减少; HR>1 表示风险增加; \*Logrank  $P<0.05$ 。

HR-hazard ratio), HR = 1: no effect; HR < 1: low risk; HR > 1: high risk; \*Logrank  $P < 0.05$ 。

对 COAD 患者而言,核心基因 TAGLN、FN1、ACTA2、DCN 表达上调对患者的 OS 和 RFS 均有显著不利影响, REG4 表达上调只对患者的 RFS 产生显著不利影响,以上 5 种核心基因表达上调的患者往往表现出较差的 OS 和/或 RFS,生存风险增加。而核心基因 ZG16 表达上调的患者往往表现出较好的 OS 和 RFS, GUCA2A、SPINK4 表达上调的患者往往表现出较好的 OS, DEFA6 表达上调的患者往往表现出较好的 RFS,以上 4 种核心基因表达上调使生存风险降低。

对 READ 患者而言,只有 3 个核心基因对患者生存时间显著性影响。如 ZG16 表达上调的患者往往表现出较好的 OS 和 RFS; SPINK4、DEFA6 表达上调的患者表现出较好的 OS。

## 2.5 核心基因与免疫细胞浸润的相关性

通过查看核心基因与肿瘤微环境中免疫细胞浸润程度之间的相关性,进一步了解核心基因在结直肠癌进展过程中的作用。提取数据,归纳总结在结直肠癌中核心基因表达水平与肿瘤纯度及肿瘤相关免疫细胞浸润程度的相关性,见表 2。

肿瘤是由恶性和非恶性细胞组成的复杂肿块。肿瘤纯度是指肿瘤组织中癌细胞所占的比例。结果显示,不论是 COAD 还是 READ,核心基因 TAGLN、FN1、ACTA2、DCN、REG4、SPINK4、ZG16 表达水平与其肿瘤纯度呈负相关;而 GUCA2A 只与 COAD 肿瘤纯度负相关,DEAF6 与结直肠癌肿瘤纯度无关。

免疫细胞浸润程度结果是在矫正肿瘤纯度之后,具有可比性。结果表明,9 种核心基因与 6 种主要免疫细胞之间存在不同程度的相关性,其中 DCN 基因在结肠腺癌中与主要免疫细胞均存在显著正相关。相较而言,CD4<sup>+</sup>T 细胞、巨噬细胞、中性粒细胞和树突细胞是被核心基因调节的主要免疫细胞群。

## 2.6 中药预测分析

通过 3 个不同数据库,经过查询和规范中药名称后,共获得中药 129 种,分别来自 Coremine Medical 数据库 29 味、BATMAN-TCM 数据库 45 味、草本植物数据库 55 味。由于 3 种数据库的来源不同,交集中药味数较少,只有半夏在 3 种数据库中同时出现,葡萄和鹿角在其中 2 个数据库中共现。因此,选择取交集对所有中药属性进行分类,以进一步了解中药特征。从功效上看,中药主要为补益药 (22)、清热解毒药 (18)、化痰止咳平喘药 (12)、祛风湿药 (11)、活血化瘀药 (9)、清热泻火药 (7)、理气药 (7);从四气属性上看,中药主要为寒性 (37)、平性 (35)、温性 (34)、凉性 (15);从五味属性上看,中药主要为甘 (60)、苦 (56)、辛 (39)、咸 (13);从归经属性上看,中药主要归肝经 (61)、肺经 (46)、脾经 (38)、肾经 (34)、胃经 (32)、大肠经 (23),以上结果见图 6。

以国家卫生健康委员会颁布及后续补充的药食同源中药名单对中药进行分类,其中 36 味药食同源中药、12 味被认为具有保健作用,两者均符合

表 2 核心基因和免疫细胞浸润程度的相关性 (Cor, P)

Table 2 Correlation between core genes and immune cell infiltration (Cor, P)

| 基因符号   | 肿瘤类型 | 肿瘤纯度                             | B 细胞                           | CD8 <sup>+</sup> T 细胞           | CD4 <sup>+</sup> T 细胞           | 巨噬细胞                            | 中性粒细胞                           | 树突状细胞                            |
|--------|------|----------------------------------|--------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|----------------------------------|
| TAGLN  | COAD | -0.366, $2.49 \times 10^{-14}$ * | -0.058, 0.245                  | 0.092, 0.065                    | 0.383, $1.79 \times 10^{-15}$ * | 0.535, $2.41 \times 10^{-31}$ * | 0.288, $4.53 \times 10^{-9}$ *  | 0.355, $2.04 \times 10^{-13}$ *  |
|        | READ | -0.282, $7.48 \times 10^{-4}$ *  | -0.057, 0.505                  | -0.154, 0.0704                  | 0.416, $3.52 \times 10^{-7}$ *  | 0.466, $7.25 \times 10^{-9}$ *  | 0.057, 0.504                    | 0.289, $5.64 \times 10^{-4}$ *   |
| FN1    | COAD | -0.332, $6.29 \times 10^{-12}$ * | -0.069, 0.168                  | 0.193, $9.40 \times 10^{-5}$ *  | 0.376, $6.39 \times 10^{-15}$ * | 0.502, $3.20 \times 10^{-27}$ * | 0.412, $7.02 \times 10^{-18}$ * | 0.447, $3.79 \times 10^{-21}$ *  |
|        | READ | -0.218, $9.65 \times 10^{-3}$ *  | -0.029, 0.731                  | 0.059, 0.489                    | 0.281, $8.14 \times 10^{-4}$ *  | 0.349, $2.49 \times 10^{-5}$ *  | 0.334, $6.36 \times 10^{-5}$ *  | 0.320, $1.23 \times 10^{-4}$ *   |
| ACTA2  | COAD | -0.362, $4.43 \times 10^{-14}$ * | 0.006, 0.904                   | 0.069, 0.165                    | 0.448, $3.28 \times 10^{-21}$ * | 0.569, $5.71 \times 10^{-36}$ * | 0.268, $4.81 \times 10^{-8}$ *  | 0.357, $1.50 \times 10^{-13}$ *  |
|        | READ | -0.271, $1.22 \times 10^{-3}$ *  | 0.011, 0.898                   | -0.136, 0.110                   | 0.465, $7.80 \times 10^{-9}$ *  | 0.423, $2.15 \times 10^{-7}$ *  | 0.049, 0.567                    | 0.363, $1.14 \times 10^{-5}$ *   |
| DCN    | COAD | -0.364, $3.19 \times 10^{-14}$ * | 0.146, $3.33 \times 10^{-3}$ * | 0.289, $2.92 \times 10^{-9}$ *  | 0.384, $1.45 \times 10^{-15}$ * | 0.646, $5.29 \times 10^{-49}$ * | 0.543, $3.40 \times 10^{-32}$ * | 0.550, $3.39 \times 10^{-33}$ *  |
|        | READ | -0.44, $5.43 \times 10^{-8}$ *   | 0.088, 0.304                   | 0.157, 0.0656                   | 0.221, $9.01 \times 10^{-3}$ *  | 0.517, $7.45 \times 10^{-11}$ * | 0.376, $5.51 \times 10^{-6}$ *  | 0.436, $8.30 \times 10^{-8}$ *   |
| REG4   | COAD | -0.332, $6.29 \times 10^{-12}$ * | -0.069, 0.168                  | 0.193, $9.40 \times 10^{-5}$ *  | 0.376, $6.39 \times 10^{-15}$ * | 0.502, $3.20 \times 10^{-27}$ * | 0.412, $7.02 \times 10^{-18}$ * | 0.447, $3.79 \times 10^{-21}$ *  |
|        | READ | -0.218, $9.65 \times 10^{-3}$ *  | -0.029, 0.731                  | 0.059, 0.489                    | 0.281, $8.14 \times 10^{-4}$ *  | 0.349, $2.49 \times 10^{-5}$ *  | 0.334, $6.36 \times 10^{-5}$ *  | 0.32, $1.23 \times 10^{-4}$ *    |
| GUCA2A | COAD | -0.099, 0.046*                   | 0.045, 0.370                   | -0.17, $5.71 \times 10^{-4}$ *  | 0.065, 0.196                    | -0.044, 0.381                   | -0.123, 0.014*                  | -0.07, 0.159                     |
|        | READ | -0.093, 0.276                    | -0.014, 0.867                  | 0.009, 0.920                    | -0.079, 0.355                   | -0.046, 0.590                   | -0.136, 0.112                   | 0.005, 0.951                     |
| SPINK4 | COAD | -0.211, $1.70 \times 10^{-5}$ *  | 0.128, $9.90 \times 10^{-3}$ * | 0.049, 0.326                    | -0.07, 0.163                    | -0.138, $5.62 \times 10^{-3}$ * | -0.062, 0.217                   | -0.003, 0.948                    |
|        | READ | -0.267, $1.45 \times 10^{-3}$ *  | 0.109, 0.202                   | 0.044, 0.610                    | -0.131, 0.124                   | -0.190, 0.025*                  | 0.014, 0.870                    | 0.077, 0.365                     |
| DEFA6  | COAD | 0.03, 0.545                      | -0.016, 0.741                  | -0.067, 0.175                   | -0.072, 0.152                   | -0.096, 0.054                   | -0.197, $7.24 \times 10^{-5}$ * | -0.156, $1.760 \times 10^{-3}$ * |
|        | READ | 0.095, 0.264                     | 0.048, 0.571                   | 0.105, 0.220                    | -0.130, 0.128                   | -0.062, 0.466                   | -0.194, 0.023*                  | 0.007, 0.939                     |
| ZG16   | COAD | -0.115, 0.021*                   | 0.091, 0.067                   | -0.142, $4.03 \times 10^{-3}$ * | 0.002, 0.974                    | -0.120, 0.015*                  | -0.115, 0.023*                  | -0.056, 0.260                    |
|        | READ | -0.265, $1.53 \times 10^{-3}$ *  | 0.082, 0.334                   | 0.032, 0.710                    | -0.12, 0.159                    | -0.160, 0.060                   | -0.036, 0.674                   | 0.066, 0.437                     |

Cor, P-相关性, P 值; \*P&lt;0.05 具有统计学意义。

Cor, P-correlation, P value; \*P&lt;0.05 is statistically significant.

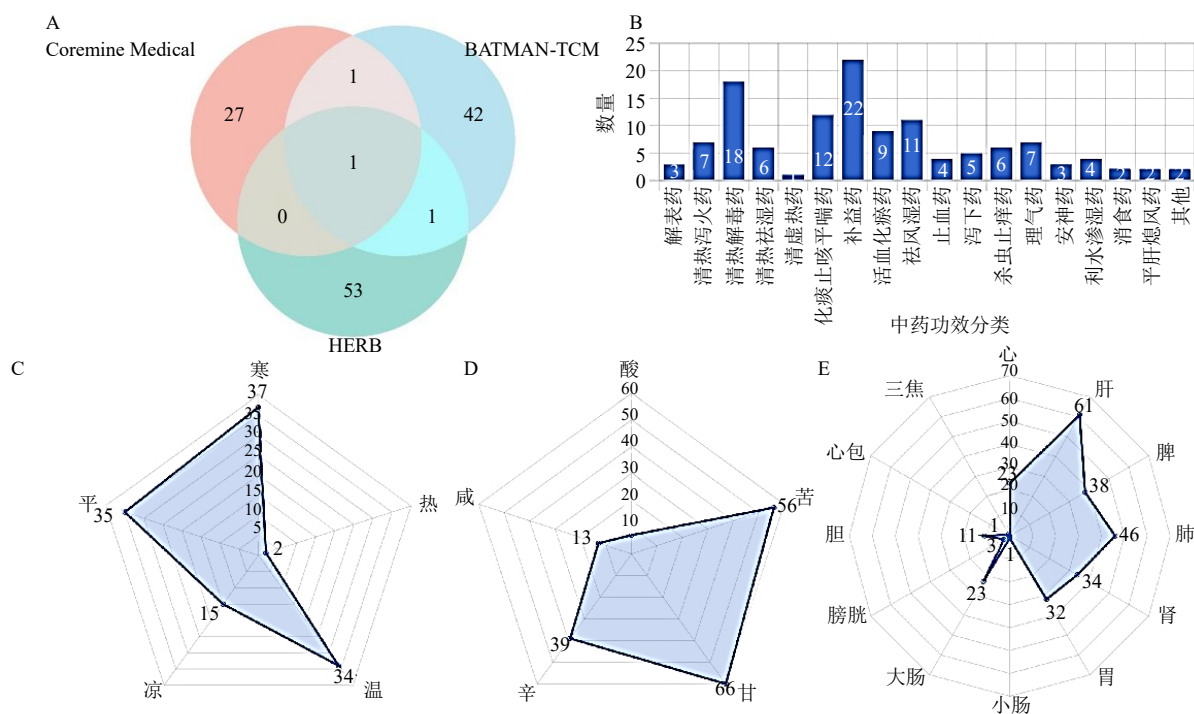


图 6 中药预测可视化

Fig. 6 Prediction visualization of traditional Chinese medicines



的有当归、韭菜籽、鳖甲、人参，27 味中药被传统中医理论认为具有毒性，结果见表 3。

2.7 中药活性成分筛选及分子对接

根据靶向基因和中药之间的 *P* 值进行度值分析，以筛选核心中药模块进行后续的活性成分筛选和分子对接。TAGLN 匹配到玉米须、葡萄、菊苣、昆布等 21 味中药；FN1 匹配到鹿角胶、肉苁蓉、鳖甲、地龙等 69 味中药；ACAT2 匹配到竹笋、丹参、桑椹、金荞麦等 107 味中药；DCN 匹

配到鹿角、枸杞、冬虫夏草等 60 味中药；GUCA2A 匹配到洋葱头、九香虫等 17 味中药；SPINK4 匹配到茯苓、茯神、向日葵籽 3 味中药；ZG16 匹配到绿豆、土茯苓等 8 味中药；REG 匹配到柠檬、山楂、酸枣仁、乳没等 32 味中药。将最终得到的中药输入 Cytoscape 软件进行中心性分析和可视化，其中葡萄、金荞麦、半夏、地龙、肉苁蓉的中心度值较高，五味核心中药进行后续分析，见图 7-A。

表 3 中药分类

Table 3 Category of traditional Chinese medicines

| 序号 | 分类   | 中药名称                                                                                                               |
|----|------|--------------------------------------------------------------------------------------------------------------------|
| 1  | 药食两用 | 玉米须、羊肾、竹笋、洋葱、茯苓、葡萄、肉苁蓉、鳖甲、昆布、菊苣、蜂蜜、当归、桑椹、绿豆、枸杞子、芡实、南瓜子、韭菜籽、大蒜、葱花/实/白、淡豆豉、百合、广枣、莱阳梨、柠檬、松子仁、羊肉、山楂、酸枣仁、人参、猪肠/胆/髓、向日葵籽 |
| 2  | 保健添加 | 金荞麦、人参、土茯苓、牛膝、贝母、川芎、鹿茸、丹参、当归、诃子、韭菜籽、鳖甲                                                                             |
| 3  | 毒性药物 | 蓖麻籽、原蚕蛾、蕲蛇、蛇莓、半夏、干蟾、马陆、獾油、蜈蚣、长春花、商陆、鼠李、藜芦、射干、望江南、鬼盖、开口箭、金鸡纳、香加皮、川乌、土茯苓、芥子、金钱白花蛇、三台红花、黄花烟草、瓜蒂、莲生桂子花                 |

将核心中药导入本草组鉴中查询化学成分，初步获取 493 个化合物，经参数标准和 Swiss ADME 数据库筛选后确定缬氨酸 (id: 6287)、鼠李糖 (id: 25310)、脯氨酸 (id: 145742)、橙花醛 (id: 643779)、蛋氨酸 (id: 84815)、甘氨酸 (id: 750) 6 种中药小分子活性成分 (表 4)。将小分子与核心中药模块中靶向目标进行对接，以分子对接结合能作为评价对接结果验证的指标，在所有对接结果中，结合能均小于-2.9 kcal/mol (1 kcal=4.2 kJ)，表明小分子与关键靶点有较好的结合能力，可以在不需要外界能量的条件下进行对接<sup>[36]</sup>，见图 7-B。将结合作用较强的前 5 位进行可视化，见图 7-C~G。

3 讨论

3.1 研究价值

结直肠癌是全球发病率稳居前 3 的癌症，早期诊断和预防是工作的重要方面。虽然胶囊内镜、粪便 DNA、影像学等技术提高了肠道肿瘤的检出率，但预防治疗仍依赖内镜下切除。然而，内镜治疗接受度普遍不高是消化道早癌预防中的棘手问题，例如韩国和日本的内镜筛查依从率分别为 30.5%<sup>[37]</sup>和 16.3%<sup>[38]</sup>，而我国城市抽检的民众依从率仅 10%~20%<sup>[39]</sup>，推广存在以下挑战<sup>[40]</sup>：(1) 民众认识不足；(2) 术前准备复杂；(3) 属于有创、侵入性治疗；(4) 介意暴露隐私部位；(5) 存在一定风险，出血、

穿孔等；(6) 需要定期复查；(7) 区域差异、医疗资源的分配等。因此，寻找替代补充的防治方法十分必要。化学药由于不良反应限制，作为预防应用受到了诸多瓶颈。而中医药运用在该方面存在如下优势：(1) 多靶点、多途径，利于改善复杂的肿瘤微环境；(2) 不良反应较少，长期作为预防用药，安全性较高；(3) 历史悠久，尤其是药食同源类中药，同时具有食物营养价值和药物治疗作用，民众认可度高；(4) 服用方便，且剂型多样；(5) 分布广泛，获取容易，价格容易接受；(6) 能减轻医疗机构负担、优化医疗资源等。因此，本研究首次聚焦结直肠癌的全程序列转化，利用生物信息学技术尝试挖掘序列转化过程的核心差异基因，预测具有疗效的中药，以优化临床用药。

3.2 靶向核心差异基因的筛选与验证

首先，选择了同时患有肠腺瘤的结直肠癌患者，分析配对样本的基因表达谱。通过两两对比 (正常黏膜 vs 肠腺瘤，肠腺瘤 vs 肠癌) 获得两阶段中重复关键差异基因 48 个。进一步通过 PPI 网络模型筛选得到 9 个核心基因，分别是 (1) TAGLN：该基因编码的蛋白质是一种对转化和形状变化敏感的肌动蛋白，充当肿瘤抑制因子，其表达丧失是细胞转化和肿瘤发展的早期事件<sup>[41]</sup>。(2) FN1：是细胞外基质的重要组成部分，参与细胞黏附和迁移

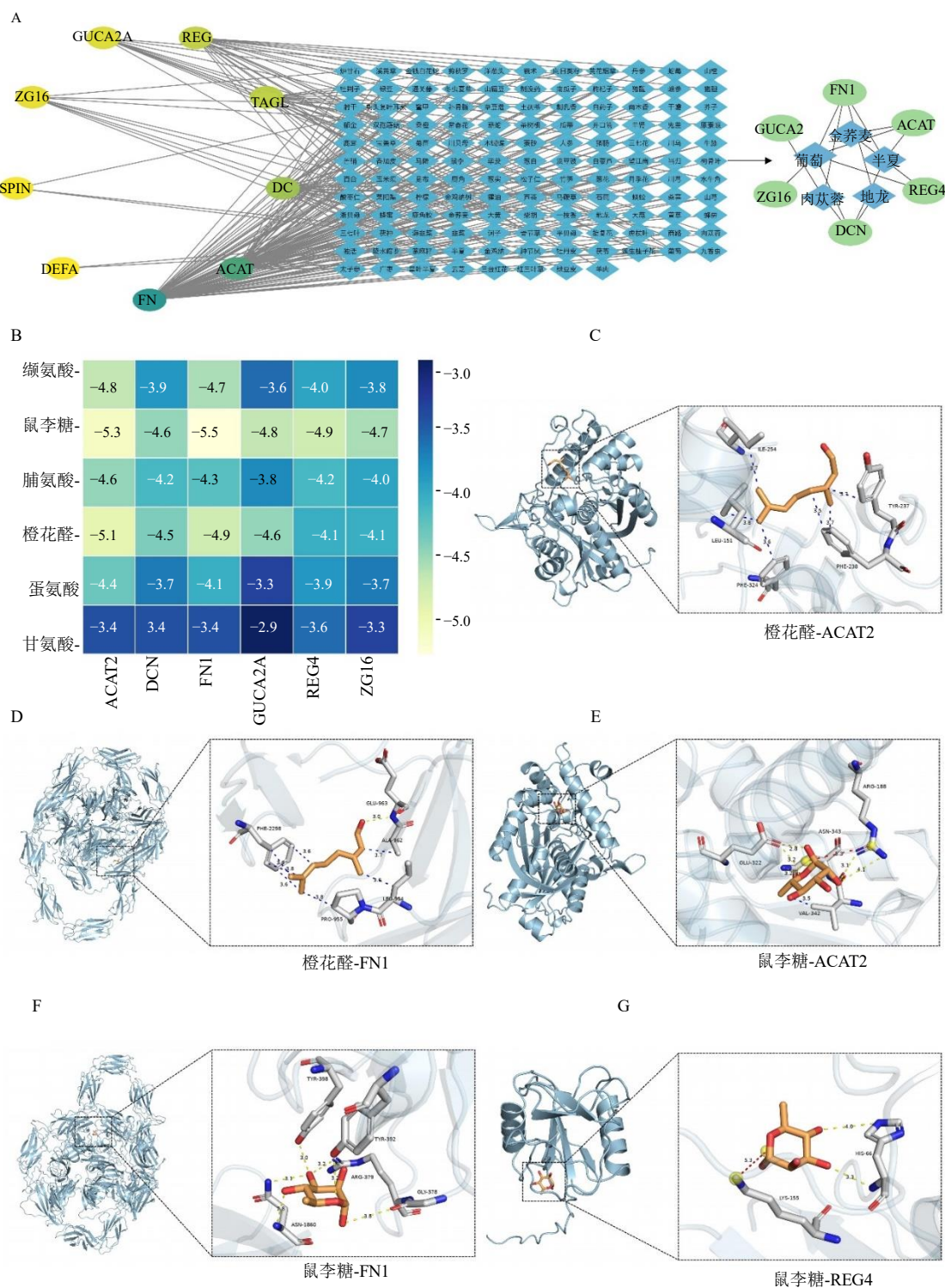


图 7 中药靶向网络 (A) 及分子对接 (B~G)

Fig. 7 Traditional Chinese medicine targeting network (A) and molecular docking (B—G)

过程, 包括伤口愈合、血液凝固、宿主防御、转移等生化过程, 在结直肠癌进展中呈动态变化, 与肿瘤形态、分期等密切相关<sup>[42]</sup>。(3) ACTA2: 参与细胞运动、结构、完整性和细胞间信号传导, 过度表达与

较差的预后相关<sup>[43]</sup>。(4) DCN: 编码一种小型富含亮氨酸的蛋白聚糖家族蛋白, 在胶原纤维组装中发挥作用, 对结直肠癌具有抑制作用<sup>[44]</sup>。(5) REG4: 对结直肠肿瘤发生至关重要, 转基因小鼠已证明

表 4 核心化合物 ADME 参数  
Table 4 ADME parameters of core compounds

| 核心化合物 | 英文名称         | PubChem_id | GI-absorbtion | Lipinski | Ghose | Veber | Egan |
|-------|--------------|------------|---------------|----------|-------|-------|------|
| 缬氨酸   | valine       | 6287       | 高             | 高        | 否     | 是     | 是    |
| 鼠李糖   | L-rhamnose   | 25310      | 高             | 高        | 否     | 是     | 是    |
| 脯氨酸   | proline      | 145742     | 高             | 高        | 否     | 是     | 是    |
| 橙花醛   | neral        | 643779     | 高             | 高        | 否     | 是     | 是    |
| 蛋氨酸   | D-methionine | 84815      | 高             | 高        | 否     | 是     | 是    |
| 甘氨酸   | glycine      | 750        | 高             | 高        | 否     | 是     | 是    |

GI-absorption-胃肠吸收; Lipinski-Lipinski 法则; Ghose-Ghose 规则; Veber-Veber 法则; Egan-Egan 规则。

GI-absorption- gastrointestinal absorption; Lipinski-Lipinski's rule; Ghose-Ghose rule; Veber-Veber rule; Egan-Egan rule.

REG4 通过 Wnt 家族分泌蛋白/ $\beta$ -连环蛋白通路 (canonical Wnt/ $\beta$ -catenin pathway) 促进癌症<sup>[45]</sup>。(6) GUCA2A: GUCA2A 在息肉转化为结直肠癌组织中起重要作用, 经 qPCR 验证其可作为结直肠癌的生物标志物之一<sup>[46]</sup>。(7) SPINK4: 主要在胃肠道和免疫系统中表达, 具有很高的结直肠癌诊断效用<sup>[47]</sup>。(8) DEFA6: 参与肠道先天免疫, 在肿瘤早期阶段发生突变, 当过表达超过正常水平 60 倍可区分个别患者的腺瘤和结肠癌<sup>[48]</sup>。(9) ZG16: 通过调控 PD-L1 表达促进结直肠癌局部免疫<sup>[49]</sup>。以上核心差异基因除了有以上文献研究的支持, 本研究还利用不同的数据库分别验证其在基因组和蛋白质组水平上的表达, 结果支持其在肿瘤中的表达差异具有统计学意义。

根据以上核心基因的主要功能, 可以将其大致分为 3 类。(1) 调控增殖相关: TAGLN、REG4、GUCA2A; (2) 细胞外基质相关: FN1、ACTA2、DCN; (3) 免疫相关: SPINK4、DEFA6、ZG16。通过功能富集分析只得到了少量的功能条目, 并没有富集到 KEGG 通路 (即无通路偏向性)。这是可以理解的, 因为本研究筛选跨度较广, 贯穿了“正常黏膜-腺瘤-腺癌”的整个序列转化, 得到的核心基因在转化过程始终呈现显著的表达差异, 它们可能与各阶段其他差异基因相互作用从而影响特定的分子生物通路。从 BP 分析结果看, 核心基因参与了对细菌存在而引发的生物反应过程, 这提示这类基因直接或间接地参与了肠道免疫; CC 分析结果表明核心基因参与细胞外基质组成; 其实, 细胞外基质和免疫系统的动态互作对揭示复杂疾病驱动机制至关重要<sup>[50]</sup>, 细胞外基质除了参与肿瘤的晚期转移, 同时其中丰富的蛋白成分及免疫活性分子对免疫系统发挥调节作用, 反之, 免疫系统可维持基

质微环境稳态、修复损伤后基质完整性, 或许这也是后 2 类核心基因在序列转化过程中的作用倾向。

基于以上功能分析, 继续进行其与免疫细胞及远期预后的相关性分析。已知结直肠癌患者免疫微环境的特征为 T、B 免疫细胞浸润数量少于健康人, 而巨噬细胞、中性粒细胞和树突状细胞在肿瘤组织中的浸润数量大于健康人<sup>[51]</sup>; 对于结直肠癌而言, T、B 细胞浸润有利于患者预后, 巨噬细胞、中性粒细胞和树突状细胞的浸润则不利于患者的预后<sup>[52]</sup>。分析结果表明核心差异基因普遍与 CD4<sup>+</sup>T 细胞、巨噬细胞、中性粒细胞、树突状细胞的分布程度显著相关。相关性分析从一定程度上验证了假设: 核心基因除了直接影响肿瘤干细胞增殖, 还能直接或间接调节免疫等微环境影响结直肠癌的发生。大部分核心基因在正常或肿瘤组织中的表达水平与生存时间的相关性之间的关系呈现非一致性, 这是因为很多基因在癌症生物学中的表达和功能取决于癌症的类型和阶段, 以抑瘤因子 TAGLN 为例, 其表达的缺失或降低会启动细胞非正常转化, 从而驱动腺瘤和腺癌的病理过程<sup>[41]</sup>, 因此较正常肠组织相比, 肠癌组织中该基因表达相对较低。而当肠癌组织中该基因表达升高时, 其可能降低了肿瘤细胞的增殖速率, 但相关研究分析显示此时 TAGLN 的高表达与肠癌细胞迁移、侵袭、血管生成和 p38 丝裂原活化蛋白激酶 (mitogen activated protein kinases, MAPK) 信号传导等密切相关, 有利于肿瘤的晚期转移, 进而对患者生存期产生不利影响<sup>[53]</sup>。总而言之, 筛选得到的核心基因不仅存在表达水平的差异, 并且对肠癌患者的免疫功能和远期预后均存在一定影响。

### 3.3 中药预测分析与验证

大肠肿瘤属于“癥瘕”“积聚”“肠溜”“肠覃”

等中医范畴,本虚标实为病机,即正气亏虚状态下,“瘀、毒、痰”等病理因素结聚发为有形实邪<sup>[54]</sup>。因此,肠癌的序列转化实际上就是由虚致实的过程。针对本虚,诸多医家责之于脾胃,认为脾胃虚寒为病因关键,虚寒导致气滞血凝痰结,久而郁结化热为息肉<sup>[55]</sup>,后期热毒深伏入络,暗伤阴液为癌<sup>[56]</sup>。而本研究筛选得到的中药,从功效上看,以补益药(陵水暗罗、韭菜子、葱实、羊肉、猪髓等)、化痰药(贝母、掌叶半夏、芥子等)、清热药(土茯苓、金荞麦、绿豆皮等)为主,符合病机演变。(1)早期重补益(无肠道实质病变,具备多项危险因素):以补脾气、壮肾阳为主,先后天同补有利于改善中焦虚寒的微环境,预防有形息肉的形成;(2)中期重清痰热(肠道腺瘤):以清热祛湿化痰为主,预防痰热胶着进展为癌;(3)后期重清热毒、通瘀络(成癌):运用清热解毒、通络祛湿、活血止血等药物。

从病位上看,该类中药以肝、肺、脾经为主,肺与大肠相表里,肺气肃降、肝主疏散,脾升胃降,体内气机与大肠传输功能密切相关。从五味特征上看,多为苦、甘、辛,苦味“能泄、能燥、能坚”,具有清热、降气、通便、燥湿、坚阴等作用;甘味“能补、能缓、能和”,能够补益和中、调和舒缓;辛味“能散、能行”,可以行气、化痰,祛湿、散结。从四气特征上看,多为寒、平、温,寒能清热、泻火、凉血、解热毒,多用于中后期;温能温中、散寒、补阳、活血、通络,用于早期或晚期;药性平和者多为药食同源类中药;极少热性药物。药物筛选得到的中药特征与另一数据挖掘抗炎、抗肿瘤中药特征的研究结果相似<sup>[57]</sup>。通过不同性味的中药组合能够在序列转化的不同阶段进行三因制宜、辨证论治,是中医药预防肿瘤等重大疾病的优势和原则。

另外,筛选得到的中药较多药食同源。而消化系统肿瘤与饮食密切相关,因此,深入挖掘此类中药来预防消化道肿瘤具有重要意义。中医肿瘤学先驱之一的周岱涵教授曾经提出:中医肿瘤食疗学是以中医脏腑经络学说理论为基础,结合现代肿瘤学和营养学,指导人们通过饮食变化治疗肿瘤的一门学科<sup>[58]</sup>。本研究挖掘出 36 味药食同源中药,按其主要功效分为补益药(人参、韭菜子、葱实、牛膝、肉苁蓉、葡萄、百合、羊肉/肾、猪髓、向日葵籽、鳖甲、桑椹、枸杞子、鹿茸、蜂蜜),清热药(金荞麦、菊苣、土茯苓、绿豆、竹笋、猪肠)、化痰药(莱

阳梨、柠檬、贝母、昆布),祛湿药(茯苓、玉米须),杀虫药(大蒜、南瓜子),解表药(葱花、葱白、淡豆豉、洋葱),行气散瘀药(广枣、山楂、丹参、川芎、当归、),温中药(芫荽),收敛药(诃子),润下药(松子仁),安神药(酸枣仁)。其中,一部分中药被作为食物香料,如葱和大蒜,二者含有丰富的大蒜素,系统性综述认为该类香料对结直肠癌细胞具有抗增殖作用,并且能抑制血管生成<sup>[59]</sup>。葱提取物能有效抑制小鼠肠道肿瘤细胞在体内外的生长<sup>[60]</sup>;来自病例对照和回顾性队列的临床研究支持食用葱属类能降低胃癌( $OR=0.54$ , 95% CI [0.43, 0.65]),剂量反应荟萃分析结果表明每天食用 20 g 葱属蔬菜(大蒜、葱、韭菜等)时,其患病风险比为  $OR=0.91$ , 95% CI [0.88, 0.94]<sup>[61]</sup>。前瞻性队列研究结果表明适量食用大蒜与降低总体结直肠癌风险显著相关( $HR=0.7$ , 95% CI [0.54, 0.91],  $P=0.007$ ),呈现 U 形剂量反应模式<sup>[62]</sup>。另一大类被作为果蔬类,如葡萄、松子仁、酸枣仁、柠檬、莱阳梨、韭菜等,是结直肠癌明确的保护性因素。该类食品含有丰富的酚酸、黄酮类、白藜芦醇、山柰酚、花青素等成分,已被广泛证实具有抑制细胞增殖、抗炎、抗氧化、调节免疫等多种生物效应<sup>[63]</sup>。

本研究筛选得到了 5 味核心中药金荞麦、半夏、葡萄、肉苁蓉、地龙,根据类药性评估筛选得到了关键的活性成分,分子对接结合能表明其与已知靶标对接良好,能够直接作用。在分子水平上阐释了潜在活性成分的药效构象和相互结合模式,进一步支持该类中药能够防治结直肠癌的作用假说。值得注意的是,筛选到了 27 味具有毒性的中药,土茯苓、半夏、芥子、金钱白花蛇、三台红花、川乌、黄花烟草、瓜蒂、莲生桂子花,这些中药多具有肝肾毒性,不能长期运用。

#### 4 结论

本研究实质上是对公共数据库的二次挖掘,存在以下不足:(1)符合要求的样本量较少,导致数据分析具有一定偏倚;(2)本研究缺少对相关参数对比的方法学研究;(3)缺乏临床样本和体内实验等直接证据的支持。但本研究严谨筛选 GSE 数据集,选择来自同一个体的配对样本进行分析,这样能够避免个体差异造成的选择偏倚。利用 2 种以上生物数据库,在不同表达水平进行表达验证,并进行预后、免疫细胞浸润等相关性分析,确认核心差异基因具有靶向治疗意义。基于此开展的靶向中药

预测分析结果可信度较高, 并利用分子对接技术进行确认, 对于早癌筛查和肿瘤防治具有一定价值。

**利益冲突** 所有作者均声明不存在利益冲突

#### 参考文献

- [1] Bray F, Laversanne M, Sung H, *et al.* Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries [J]. *CA Cancer J Clin*, 2024, 74(3): 229-263.
- [2] Keum N, Giovannucci E. Global burden of colorectal cancer: Emerging trends, risk factors and prevention strategies [J]. *Nat Rev Gastroenterol Hepatol*, 2019, 16(12): 713-732.
- [3] Leslie A, Carey F A, Pratt N R, *et al.* The colorectal adenoma-carcinoma sequence [J]. *Br J Surg*, 2002, 89(7): 845-860.
- [4] van der Meulen M P, Lansdorp-Vogelaar I, Goede S L, *et al.* Colorectal cancer: Cost-effectiveness of colonoscopy versus CT colonography screening with participation rates and costs [J]. *Radiology*, 2018, 287(3): 901-911.
- [5] Gubbiotti A, Spadaccini M, Badalamenti M, *et al.* Key factors for improving adenoma detection rate [J]. *Expert Rev Gastroenterol Hepatol*, 2022, 16(9): 819-833.
- [6] Albéniz E, Montori S, Rodríguez de Santiago E, *et al.* Preventing postendoscopic mucosal resection bleeding of large nonpedunculated colorectal lesions [J]. *Am J Gastroenterol*, 2022, 117(7): 1080-1088.
- [7] Lansdorp-Vogelaar I, Knudsen A B, Brenner H. Cost-effectiveness of colorectal cancer screening [J]. *Epidemiol Rev*, 2011, 33(1): 88-100.
- [8] Rotermund C, Djinbachian R, Taghiakbari M, *et al.* Recurrence rates after endoscopic resection of large colorectal polyps: A systematic review and meta-analysis [J]. *World J Gastroenterol*, 2022, 28(29): 4007-4018.
- [9] Veettil S K, Lim K G, Ching S M, *et al.* Effects of aspirin and non-aspirin nonsteroidal anti-inflammatory drugs on the incidence of recurrent colorectal adenomas: A systematic review with meta-analysis and trial sequential analysis of randomized clinical trials [J]. *BMC Cancer*, 2017, 17(1): 763.
- [10] Ni W J, Liu T, Liu Y J, *et al.* Sijunzi decoction granules in the prevention and treatment of recurrence of colorectal adenoma: Study protocol for a multicenter, randomized, double-blind, placebo-controlled trial [J]. *Front Pharmacol*, 2023, 14: 1175811.
- [11] Sui H, Zhang L, Gu K J, *et al.* YYFZBJS ameliorates colorectal cancer progression in Apc<sup>Min/+</sup> mice by remodeling gut microbiota and inhibiting regulatory T-cell generation [J]. *Cell Commun Signal*, 2020, 18(1): 113.
- [12] 张转红, 刘婷, 高飞云, 等. 中药抑制结肠癌血管生成的作用机制研究进展 [J]. 中草药, 2023, 54(3): 948-955.
- [13] 曾庆昇, 孙轶, 杨红杰, 等. CRC 候选关键基因及潜在疗效中药生物信息学分析 [J]. 中南医学科学杂志, 2023, 51(1): 19-23.
- [14] 刘洋, 曹赛雅, 冯月娇, 等. 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 [J]. 中草药, 2023, 54(19): 6386-6399.
- [15] 边杰, 王涛, 叶长春, 等. 高通量 RNA 测序技术分析结肠腺瘤转变为结肠癌的关键基因 [J]. 中国现代普通外科进展, 2024, 27(4): 286-291.
- [16] Pan Z F, He Y, Zhu W J, *et al.* A dynamic transcription factor signature along the colorectal adenoma-carcinoma sequence in patients with co-occurrent adenoma and carcinoma [J]. *Front Oncol*, 2021, 11: 597447.
- [17] Barrett T, Wilhite S E, Ledoux P, *et al.* NCBI GEO: Archive for functional genomics data sets: Update [J]. *Nucleic Acids Res*, 2013, 41(Database issue): D991-D995.
- [18] Shklar M, Strichman-Almashanu L, Shmueli O, *et al.* GeneTide: Terra Incognita Discovery Endeavor: A new transcriptome focused member of the GeneCards/GeneNote suite of databases [J]. *Nucleic Acids Res*, 2005, 33(Database issue): D556-D561.
- [19] Szklarczyk D, Kirsch R, Koutrouli M, *et al.* The STRING database in 2023: Protein-protein association networks and functional enrichment analyses for any sequenced genome of interest [J]. *Nucleic Acids Res*, 2023, 51(D1): D638-D646.
- [20] Shannon P, Markiel A, Ozier O, *et al.* Cytoscape: A software environment for integrated models of biomolecular interaction networks [J]. *Genome Res*, 2003, 13(11): 2498-2504.
- [21] Zhou Y Y, Zhou B, Pache L, *et al.* Metascape provides a biologist-oriented resource for the analysis of systems-level datasets [J]. *Nat Commun*, 2019, 10(1): 1523.
- [22] 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.
- [23] Blum A, Wang P, Zenklusen J C. SnapShot: TCGA-analyzed tumors [J]. *Cell*, 2018, 173(2): 530.
- [24] GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans [J]. *Science*, 2015, 348(6235): 648-660.
- [25] Uhlen M, Oksvold P, Fagerberg L, *et al.* Towards a knowledge-based Human Protein Atlas [J]. *Nat Biotechnol*, 2010, 28(12): 1248-1250.
- [26] Long S D, Ji S S, Xiao K M, *et al.* Prognostic and immunological value of LTB4R in pan-cancer [J]. *Math Biosci Eng*, 2021, 18(6): 9336-9356.
- [27] Li T W, Fu J X, Zeng Z X, *et al.* TIMER2.0 for analysis of tumor-infiltrating immune cells [J]. *Nucleic Acids Res*, 2020, 48(W1): W509-W514.
- [28] Fang S S, Dong L, Liu L, *et al.* HERB: A high-throughput experiment- and reference-guided database of traditional Chinese medicine [J]. *Nucleic Acids Res*, 2021, 49(D1): D1197-D1206.
- [29] 尹倩, 曾剑锋, 蒋力生. 中医食疗养生应用特点分析 [J]. 中华中医药杂志, 2021, 36(4): 2378-2380.
- [30] Ru J L, Li P, Wang J N, *et al.* TCMSP: A database of systems pharmacology for drug discovery from herbal medicines [J]. *J Cheminform*, 2014, 6: 13.
- [31] Liu H, Wang J N, Zhou W, *et al.* Systems approaches and polypharmacology for drug discovery from herbal



- medicines: An example using licorice [J]. *J Ethnopharmacol*, 2013, 146(3): 773-793.
- [32] Falaguera M J, Mestres J. Identification of the core chemical structure in SureChEMBL patents [J]. *J Chem Inf Model*, 2021, 61(5): 2241-2247.
- [33] Berman H M, Westbrook J, Feng Z, *et al.* The protein data bank [J]. *Nucleic Acids Res*, 2000, 28(1): 235-242.
- [34] Morris G M, Huey R, Lindstrom W, *et al.* AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility [J]. *J Comput Chem*, 2009, 30(16): 2785-2791.
- [35] Trott O, Olson A J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading [J]. *J Comput Chem*, 2010, 31(2): 455-461.
- [36] Gaillard T. Evaluation of AutoDock and AutoDock vina on the CASF-2013 benchmark [J]. *J Chem Inf Model*, 2018, 58(8): 1697-1706.
- [37] Park B, Her E Y, Lee K, *et al.* Overview of the national cancer screening program for colorectal cancer in Korea over 14 years (2004-2017) [J]. *Cancer Res Treat*, 2023, 55(3): 910-917.
- [38] Hamashima C, Goto R. Potential capacity of endoscopic screening for gastric cancer in Japan [J]. *Cancer Sci*, 2017, 108(1): 101-107.
- [39] 李晴雨, 杨雷, 张希, 等. 北京市城区上消化道癌高危人群内镜筛查依从性及其影响因素分析 [J]. *中国公共卫生*, 2022, 38(11): 1445-1449.
- [40] Matsuda T. Expectations for and challenges in population-based endoscopic gastric and colorectal cancer screening [J]. *Dig Endosc*, 2022, 34(Suppl 2): 15-19.
- [41] Zhou Y, Bian S H, Zhou X, *et al.* Single-cell multiomics sequencing reveals prevalent genomic alterations in tumor stromal cells of human colorectal cancer [J]. *Cancer Cell*, 2020, 38(6): 818-828.
- [42] Wang J N, Li R B, Li M Y, *et al.* Fibronectin and colorectal cancer: Signaling pathways and clinical implications [J]. *J Recept Signal Transduct Res*, 2021, 41(4): 313-320.
- [43] Kotsiliti E. Origin of CAFs in colorectal cancer [J]. *Nat Rev Gastroenterol Hepatol*, 2022, 19(2): 79.
- [44] Valdeolivas A, Amberg B, Giroud N, *et al.* Profiling the heterogeneity of colorectal cancer consensus molecular subtypes using spatial transcriptomics [J]. *NPJ Precis Oncol*, 2024, 8(1): 10.
- [45] Hwang J H, Yoon J, Cho Y H, *et al.* A mutant KRAS-induced factor REG4 promotes cancer stem cell properties via Wnt/ $\beta$ -catenin signaling [J]. *Int J Cancer*, 2020, 146(10): 2877-2890.
- [46] Hosseini S T, Nemat F. Identification of GUCA2A and COL3A1 as prognostic biomarkers in colorectal cancer by integrating analysis of RNA-Seq data and qRT-PCR validation [J]. *Sci Rep*, 2023, 13(1): 17086.
- [47] Liao C C, Wang Q, An J X, *et al.* SPINKs in tumors: Potential therapeutic targets [J]. *Front Oncol*, 2022, 12: 833741.
- [48] Radeva M Y, Jahns F, Wilhelm A, *et al.* Defensin alpha 6 (DEFA 6) overexpression threshold of over 60 fold can distinguish between adenoma and fully blown colon carcinoma in individual patients [J]. *BMC Cancer*, 2010, 10: 588.
- [49] Meng H, Ding Y, Liu E J, *et al.* ZG16 regulates PD-L1 expression and promotes local immunity in colon cancer [J]. *Transl Oncol*, 2021, 14(2): 101003.
- [50] Ganassi M, Zammit P S. Involvement of muscle satellite cell dysfunction in neuromuscular disorders: Expanding the portfolio of satellite cell-opathies [J]. *Eur J Transl Myol*, 2022, 32(1): 10064.
- [51] 吴林峰, 孙永吉, 叶乐乐, 等. 结直肠癌组织浸润免疫细胞组成及特征 [J]. *温州医科大学学报*, 2020, 50(6): 454-460.
- [52] Mizuno R, Kawada K, Itatani Y, *et al.* The role of tumor-associated neutrophils in colorectal cancer [J]. *Int J Mol Sci*, 2019, 20(3): 529.
- [53] Pini A, Garella R, Idrizaj E, *et al.* Glucagon-like peptide 2 counteracts the mucosal damage and the neuropathy induced by chronic treatment with cisplatin in the mouse gastric fundus [J]. *Neurogastroenterol Motil*, 2016, 28(2): 206-216.
- [54] 宋卓, 杨宇飞, 孙凌云, 等. 胃肠恶性肿瘤以脾虚为核心病机的理论与实践探讨 [J]. *世界中医药*, 2021, 16(9): 1357-1363.
- [55] 楚永庆, 陈霞, 李力, 等. 160 例大肠息肉患者中医证型与中医体质分布规律研究 [J]. *中医杂志*, 2021, 62(11): 966-970.
- [56] 龙思丹. 基于风险分层肠腺瘤患者舌象特征与 LTB4/BLT1 轴的相关研究 [D]. 北京: 北京中医药大学, 2021.
- [57] 陈美池, 谢虹亭, 龙思丹, 等. 基于数据挖掘的抗炎抗肿瘤中药药性特征分析 [J]. *上海中医药杂志*, 2023, 57(4): 44-50.
- [58] 李穗晖, 周瑞生, 沈曼娜. 略谈周岱翰教授对中医肿瘤食疗学贡献 [J]. *中医肿瘤学杂志*, 2019, 1(2): 66-68.
- [59] Hossain M S, Kader M A, Goh K W, *et al.* Herb and spices in colorectal cancer prevention and treatment: A narrative review [J]. *Front Pharmacol*, 2022, 13: 865801.
- [60] Arulselvan P, Wen C C, Lan C W, *et al.* Dietary administration of scallion extract effectively inhibits colorectal tumor growth: Cellular and molecular mechanisms in mice [J]. *PLoS One*, 2012, 7(9): e44658.
- [61] Zhou Y, Zhuang W, Hu W, *et al.* Consumption of large amounts of *Allium* vegetables reduces risk for gastric cancer in a meta-analysis [J]. *Gastroenterology*, 2011, 141(1): 80-89.
- [62] Jiang Z Z, Chen H L, Li M, *et al.* Garlic consumption and colorectal cancer risk in US adults: A large prospective cohort study [J]. *Front Nutr*, 2023, 10: 1300330.
- [63] Alzate-Yepes T, Pérez-Palacio L, Martínez E, *et al.* Mechanisms of action of fruit and vegetable phytochemicals in colorectal cancer prevention [J]. *Molecules*, 2023, 28(11): 4322.

[责任编辑 潘明佳]