

# $\alpha 1$ -I型和 $\alpha 1$ -XIII型胶原基因可作为评估结直肠腺瘤恶变风险的生物标志物

孙中尚,高志颖,潘峰,王宏刚,潘振国\*

南京医科大学附属淮安第一医院 消化内科,江苏 淮安 223300

**【摘要】** 目的 应用生物信息学方法探究结直肠腺瘤-腺癌序列中的关键基因,筛选出可用于评估腺瘤恶变风险的标志物。方法 搜索 GEO 数据库中结直肠腺瘤与腺癌相关数据集,筛选出关键差异基因,并对关键差异基因行基因本体(GO)功能和京都基因与基因组百科全书(KEGG)信号通路富集分析,通过构建蛋白质互作网络(PPI)筛选出主效基因,利用 GEPIA 数据库绘制生存曲线分析主效基因与结直肠腺癌预后的关系。结果 在 GSE50115 和 GSE46157 两个数据集中,筛选出 42 个关键差异基因。GO 功能分析表明,关键差异基因主要富集在组织细胞外基质、胶原构成的细胞外基质和细胞外基质结构组成等功能,KEGG 信号通路分析显示主要富集于细胞外基质受体相互作用。PPI 共筛选出 6 个主效基因,其中  $\alpha 1$ -I 型胶原基因(COL1A1)和  $\alpha 1$ -XIII 型胶原基因(COL12A1)表达量与结直肠腺癌患者预后呈负相关。结论 COL1A1 和 COL12A1 在结直肠腺瘤-腺癌序列中发挥重要作用,可作为判断腺瘤恶变风险的标志物。

**【关键词】** 结直肠腺瘤-腺癌序列; 结直肠腺瘤恶变风险; 生物信息分析

## COL1A1 and COL12A1 can be used as biomarkers to assess the risk of malignancy in colorectal adenomas

Sun Zhongshang, Gao Zhiying, Pan Feng, Wang Honggang, Pan Zhenguo\*

Department of Gastroenterology, the Affiliated Huai'an First People's Hospital of Nanjing Medical University, Huai'an 223300, Jiangsu, China

\*Corresponding author: Pan Zhenguo, E-mail: panzhenguo2019@163.com

**【Abstract】 Objective** To apply a bioinformatics approach to explore key genes in colorectal adenoma-adenocarcinoma sequences and screen for markers that can be used to assess the risk of adenoma malignancy. **Method** We searched the GEO database for colorectal adenoma-adenocarcinoma related datasets, screened out key differential genes, performed gene ontology (GO) function and Kyoto Encyclopedia of Genomes(KEGG) signaling pathway enrichment analysis of key differential genes, identified master-effect genes by constructing protein interaction network (PPI) network, and analyzed master-effect genes by plotting survival curves using GEPIA database. The relationship with prognosis of colorectal adenocarcinoma was analyzed by drawing survival curves using GEPIA database. **Result** The results showed that 42 key differential genes were screened in the GSE50115 and GSE46157 datasets, and the GO functional analysis showed that the key differential genes were mainly enriched in the functions of tissue extracellular matrix, extracellular matrix composed of collagen and extracellular matrix structural composition, and the KEGG signaling pathway analysis showed that they were mainly enriched in extracellular matrix receptor interactions. A total of 6 key genes were screened by PPI, among which the expression of Collagen alpha-1 (I) chain (COL1A1) and collagen type XIII alpha 1 chain (COL12A1) was negatively correlated with the prognosis of colorectal adenocarcinoma patients. **Conclusion** COL1A1 and COL12A1 play an important role in the colorectal adenoma-adenocarcinoma sequence and can be used as markers to determine the risk of adenoma malignancy.

**【Key words】** Colorectal adenoma-adenocarcinoma sequence; Colorectal adenoma malignancy risk; Bioinformatic analysis

基金项目:江苏省老年健康科研项目(LKM2023044)

\*通信作者:潘振国, E-mail: panzhenguo2019@163.com

在全球范围内,结直肠癌的发病率和死亡率均逐年升高。在欧美地区,由于饮食和环境等因素,其结直肠癌发病率远高于其他地区<sup>[1]</sup>。近年来,随着我国经济快速发展,人民生活水平提高,饮食结构逐渐西化,结直肠癌的发病率和死亡率呈明显上升趋势,这严重危害了我国居民的生命健康<sup>[2]</sup>。对人群行大规模肠镜筛查是降低进展期结直肠癌发病率的有效手段,同时很多结直肠腺瘤也被检出。

结直肠腺瘤是起源于大肠黏膜腺上皮的良性肿瘤,病理上可分为管状腺瘤、绒毛状腺瘤和绒毛管状腺瘤<sup>[3]</sup>。有大约85%的结直肠癌是通过腺瘤-腺癌序列这条途径发生的,其中腺瘤这一癌前病变中结直肠腺瘤性息肉病基因(adenomatous polytosis coli, APC)、抑癌基因P53、Kirsten大鼠肉瘤病毒癌基因同源物(Kirsten ratsarcoma viral oncogene homolog, KRAS)、肿瘤抑制基因(mothers against decapentaplegic homolog 4, SMAD4)等基因发生突变腺瘤转变为腺癌的关键因素,该过程通常需要5~10年<sup>[4-5]</sup>。目前,结直肠癌发生发展存在多种分子机制,但被普遍接受的观点是,结直肠癌是源于体内遗传和表观遗传学通路的改变,包括磷酸酰肌醇-3-激酶(phosphatidylinositol 3 kinase, PI3K)和转化生长因子- $\beta$  (transforming growth factor  $\beta$ , TGF- $\beta$ )等信号通路<sup>[6]</sup>。

依据结直肠腺瘤发展至腺癌的风险高低,可将其分为低危和高危腺瘤。其中腺瘤体积是判断腺瘤危险性的重要指标,一般将腺瘤长径 $\geq 10$  mm定义为高危,若腺瘤长径 $< 10$  mm但病理发现存在绒毛成分超过25%或高级别上皮内瘤变,也将被视为高危腺瘤<sup>[3]</sup>。目前,内镜下切除是治疗高危结直肠腺瘤的一线治疗方案,但若病理发现存在进展期癌变或内镜下切除困难等因素,则需要进一步手术切除,而对于低危腺瘤,则推荐定期随访内镜检查<sup>[7]</sup>,因此,对腺瘤恶变风险的精确判断是提供最佳治疗方案的前提条件。虽然高危腺瘤具有明确形态学特征,但仍有大量病变特征模糊,容易导致病理诊断发生差错。若能够发现可区分结直肠腺瘤与腺癌的特定生物标志物,将会显著提高腺瘤与腺癌的诊断水平,从而为腺瘤患者提供最优的诊疗计划。

RNA测序技术可以在全转录组水平分析样本之间的表达差异,根据测序结果,可以通过生物信

息学方法开展功能富集分析,在众多差异表达基因中筛选出关键差异基因和信号通路。因此,利用生物信息学方法分析结直肠腺瘤和腺癌之间的表达谱差异,将对深入探究结直肠腺瘤-腺癌序列提供帮助,但是测序结果容易受限于测序技术、芯片和样本之间的差异,而使误差较大。因此,本研究旨在通过降低不同芯片、方法之间的差异,以探究结直肠腺瘤-腺癌序列中的关键基因和信号通路。我们提取了美国国立生物技术信息中心的基因表达综合数据库(gene expression omnibus database, GEO数据库)中GSE50115和GSE41657两个数据集,对比分析了88个结直肠腺瘤样本和34个结直肠腺癌样本,筛选出6个主效基因,分析出 $\alpha 1$ -I型胶原基因(collagen, type I,  $\alpha 1$ , COL1A1)和 $\alpha 1$ -XII型胶原基因(collagen type XII,  $\alpha 1$ , COL12A1)在结直肠腺瘤-腺癌序列中发挥重要作用,并且有助于在分子水平判断结直肠腺瘤恶变的风险。

## 1 资料与方法

1.1 数据资料来源 从GEO数据库(<http://www.ncbi.nlm.nih.gov/geo>)中检索“colorectal adenoma”“colorectal adenocarcinoma”,下载相关数据用于后续研究。筛选标准:①样本取自人结直肠组织;②以结直肠腺瘤组织作为参照组,结直肠腺癌组织作为试验组;③腺瘤组与腺癌组样本数均大于5例,基于上述标准,我们选取了GSE50115和GSE41657两个数据集。GSE50115平台号:Agilent-014850 Whole Human Genome Microarray 4x44K G4112F (Probe Name version),包括37个结直肠腺瘤样本和9个结直肠腺癌样本。GSE41657平台号:GPL6480 Agilent-014850 Whole Human Genome Microarray 4x44K G4112F (Probe Name version),包含51个结直肠腺瘤样本和25个结直肠腺癌样本。

1.2 筛选差异表达基因 (differentially expressed genes, DEGS) 应用GEO2R工具对GSE50115和GSE41657两个数据集中结直肠腺瘤和结直肠腺癌组织样本差异基因进行筛选,筛选标准设定 $|\log_2(FC)| > 1.5$ 且 $P_{adj} < 0.05$ ,设 $\log_2(FC) > 1.5$ 为上调基因, $\log_2(FC) < -1.5$ 为下调基因。通过R语言(3.6.3版本)的ggplot2 [3.3.3版本]绘制火山图和韦恩图。利用R语言的ComplexHeatmap包[2.2.0版本]绘制表达热图。

1.3 功能富集分析 应用R语言(3.6.3版本)的

ggplot2包[3.3.3版本](用于可视化),clusterProfiler包[3.14.3版本](用于富集分析),对两个数据集行基因本体数据库(Gene Ontology,GO)分别分析生物过程(biological process,BP)、细胞成分(cell components,CC)和分子功能(molecular function, MF),京都基因与基因百科全书(Kyoto Encyclopedia of Genes and Genomes,KEGG)用于通路富集分析。

1.4 蛋白质互作网络分析 应用STRING数据库(<https://string-db.org/>)(版本11.5)对两个数据集所筛选出的共同差异基因行蛋白质互作网络分析(protein-protein interaction network analysis,PPI分析),应用Cytoscape软件(版本3.8.1)对PPI结果行可视化处理,使用该软件中的CytoHubba插件按照连接度高低,筛选出关键基因用于后续分析。

1.5 分析关键基因与结直肠癌预后的关系 基因表达谱数据动态分析(gene expression profiling interactive analysis,GEPIA)数据库分别分析关键基因与结直肠癌患者预后的关系,方法参数为无病生存(disease-free survival,DFS)期,设定高低组间临界值均为50%,Log-rank检验法用于两组患者DFS的比较, $P<0.05$ 为差异有统计学意义。

1.6 统计学分析 STRING数据库参数设置为相互作用分数为0.400(medium confidence)。Cytoscape

软件中的CytoHubba插件设置为采取MCC法,取得分最高的6个基因作为关键基因。采用R语言的Enrich GO函数和Enrich KEGG函数进行GO和KEGG功能分析,参数设置:校正 $P$ 值 $<0.05$ , $Q$ 值 $<0.05$ 。所有统计分析均采用 $P<0.05$ 为差异有统计学意义。

## 2 结果

2.1 筛选关键差异基因 在对GSE50115和GSE41657两个微阵列数据集分别做差异基因分析。将结直肠腺瘤设为参照组,结直肠癌设为实验组,以 $|\log_2(FC)|>1.5$ 且校正 $P<0.05$ 为筛选临界值,在GSE50115数据集中,满足条件的基因有249个,其中上调表达基因 $[\log_2(FC)$ 数值为正数]有149个,下调表达基因 $[\log_2(FC)$ 数值为负数]有100个,在GSE41657数据集中,达到临界值的基因有118个,其中上调表达69个,下调表达49个。两个数据集中共同表达的差异基因上调的有22个,下调则有20个。两组数据集的差异基因以火山图展示,以韦恩图展示两个数据集的关键差异基因,两个数据集中关键差异基因表达量以热图展示(图1)。

2.2 功能富集分析 对两个数据集所筛选出的

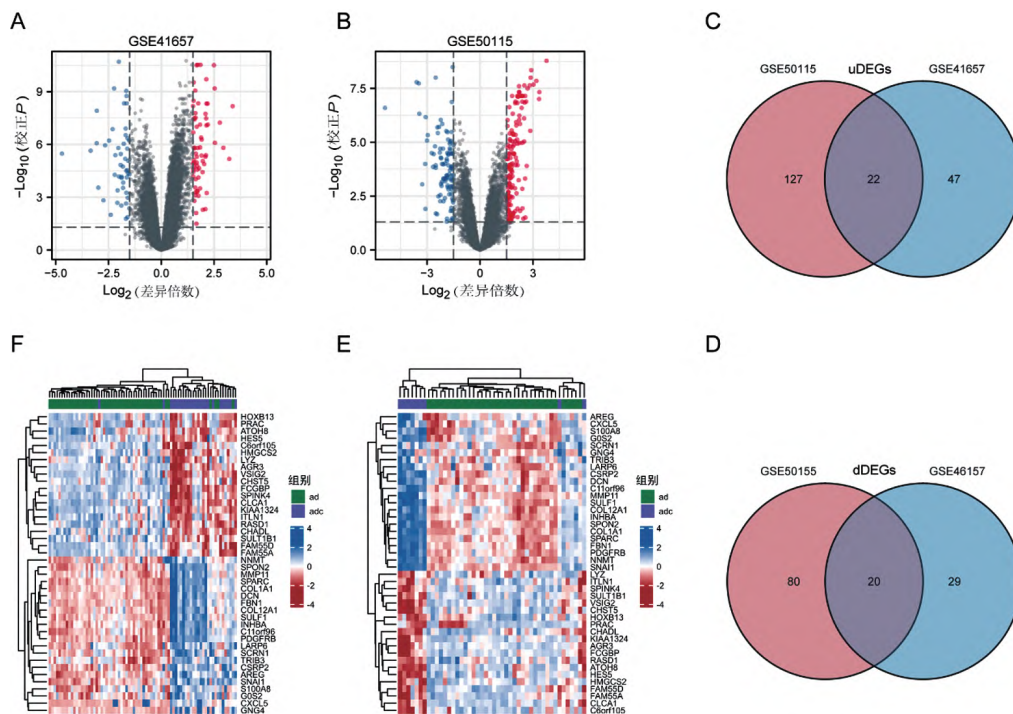


图1 筛选 GSE50115 和 GSE41657 两个数据集的关键差异基因

注:A、B分别展示基因表达水平及分布火山图;C、D分别以韦恩图展示两个数据集中关键差异基因;E、F火山图展示关键差异基因的表达水平。uDEGs为上调表达的差异基因;dDEGs为下调表达的差异基因;ad为结直肠腺瘤;adc为结直肠癌。

关键差异基因行 GO 功能和 KEGG 信号通路富集分析,GO 功能注释结果显示,关键差异基因生物学过程主要富集在细胞外结构组织和细胞外基质组织,差异基因的细胞成分主要富集在含胶原蛋白的细胞外基质和内质网腔,差异基因的分子功能主要富集在内肽酶活性、受体配体活性和细胞外基质结构成分(图 2)。KEGG 信号通路分析显示,差异基因主要富集于细胞外基质(extracellular matrix,ECM)受体相互作用通路、蛋白质消化和吸收通路以及 TGF-β 信号通路中(图 3)。关键差异

基因功能和信号通路富集分析见表 1。

2.3 蛋白质互作网络构建与主效基因筛选 利用 STRING 数据库对两个数据集的关键基因进行 PPI 网络构建,通过 Cytoscape 软件行可视化处理(图 4)。利用 Cytoscape 软件中 CytoHubba 插件评估 PPI 中的连接度,并据此筛选出 SPARC、COL1A1、COL12A1、PDGFRB、FBN1 和 DCN 共计 6 个主效基因。

2.4 主效基因与结直肠癌预后的关系 利用 GEPIA 数据库分别分析筛选出的 6 个主效基因与结直肠

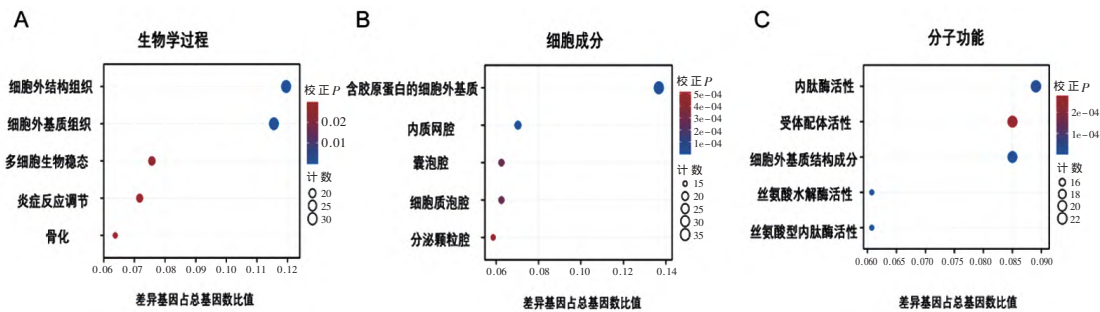


图 2 差异基因的 GO 富集功能分析

注:A 图显示生物学过程主要为细胞外结构组织、细胞外基质组织、多细胞生物稳态、炎症反应调节、骨化。B 图显示细胞成分主要为含胶原蛋白的细胞外基质、内质网腔、囊泡腔、细胞质泡腔、分泌颗粒腔。C 图显示分子功能主要为内肽酶活性、受体配体活性、细胞外基质结构成分、丝氨酸水解酶活性、丝氨酸型内肽酶活性。

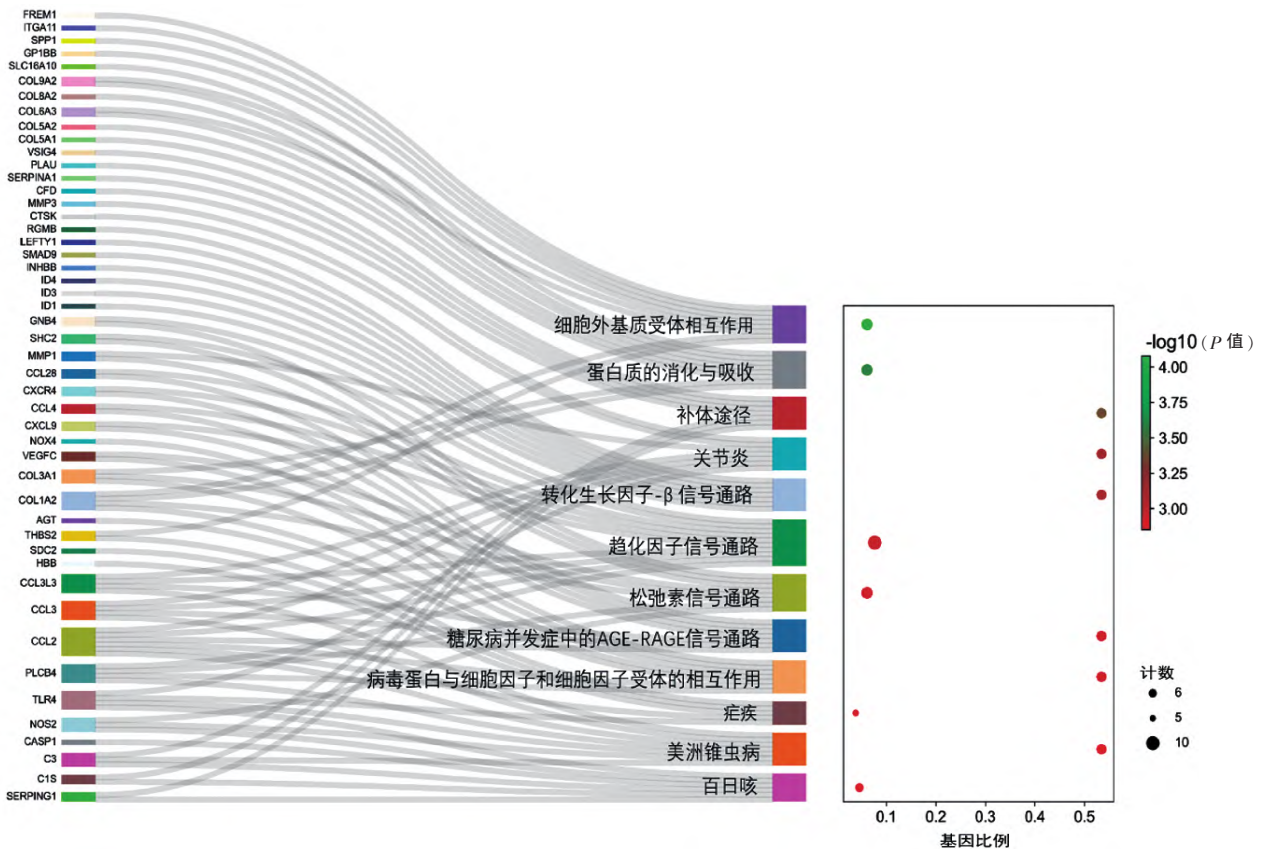


图 3 差异基因的 KEGG 信号通路分析

表1 关键差异基因功能和信号通路分析数据结果

| GO 功能注释 | ID         | 名称                     | 基因比值   | P 值          | 校正 P 值       | G 值          |
|---------|------------|------------------------|--------|--------------|--------------|--------------|
| BP      | GO:1903034 | 炎症反应调节                 | 9/251  | 0.000 726 54 | 0.072 710 4  | 0.065 582 17 |
| BP      | GO:0043062 | 细胞外结构组织                | 30/251 | 9.541E-14    | 1.4765E-10   | 1.3317E-10   |
| BP      | GO:0030198 | 细胞外基质组织                | 29/251 | 1.8257E-14   | 5.6504E-11   | 5.0965E-11   |
| BP      | GO:0048871 | 多细胞生物稳态                | 19/251 | 3.2931E-05   | 0.025 480 62 | 0.022 982 6  |
| BP      | GO:0001503 | 骨化                     | 16/251 | 0.000 103 07 | 0.029 751 89 | 0.026 835 13 |
| CC      | GO:0062023 | 含胶原蛋白的细胞外基质            | 35/256 | 5.9016E-19   | 1.5698E-16   | 1.3356E-16   |
| CC      | GO:0005788 | 内质网腔                   | 18/256 | 1.3037E-07   | 1.0113E-05   | 8.6038E-06   |
| CC      | GO:0060205 | 细胞质泡腔                  | 16/256 | 9.5883E-06   | 0.000 264 62 | 0.000 225 14 |
| CC      | GO:0031983 | 囊泡腔                    | 16/256 | 9.948E-06    | 0.000 264 62 | 0.000 225 14 |
| CC      | GO:0034774 | 分泌颗粒腔                  | 15/256 | 2.1229E-05   | 0.000 513 37 | 0.000 436 78 |
| MF      | GO:0004175 | 内肽酶活性                  | 22/247 | 1.5336E-07   | 1.3863E-05   | 1.243E-05    |
| MF      | GO:0005201 | 细胞外基质结构成分              | 21/247 | 1.1757E-14   | 5.314E-12    | 4.7645E-12   |
| MF      | GO:0048018 | 受体配体活性                 | 21/247 | 4.3445E-06   | 0.00028053   | 0.00025152   |
| MF      | GO:0004252 | 丝氨酸型内肽酶活性              | 15/247 | 7.2115E-09   | 1.6298E-06   | 1.4613E-06   |
| MF      | GO:0017171 | 丝氨酸水解酶活性               | 15/247 | 5.5136E-08   | 6.2304E-06   | 5.5862E-06   |
| KEGG    | hsa04512   | 细胞外基质受体相互作用            | 8/131  | 8.3722E-05   | 0.017 079 28 | 0.015 246 21 |
| KEGG    | hsa04974   | 蛋白质的消化与吸收              | 8/131  | 0.000 252 74 | 0.022 749 32 | 0.020 307 7  |
| KEGG    | hsa04926   | 松弛素信号通路                | 8/131  | 0.001 140 2  | 0.022 749 32 | 0.020 307 7  |
| KEGG    | hsa04610   | 补体途径                   | 7/131  | 0.000 434 32 | 0.022 749 32 | 0.020 307 7  |
| KEGG    | hsa05323   | 类风湿性关节炎                | 7/131  | 0.000 748 93 | 0.022 749 32 | 0.020 307 7  |
| KEGG    | hsa04350   | 转化生长因子- $\beta$ 信号通路   | 7/131  | 0.000 798 43 | 0.022 749 32 | 0.020 307 7  |
| KEGG    | hsa04061   | 病毒蛋白与细胞因子和细胞因子受体的相互作用  | 7/131  | 0.001 152 37 | 0.022 749 32 | 0.020 307 7  |
| KEGG    | hsa04933   | 糖尿病并发症中的 AGE-RAGE 信号通路 | 7/131  | 0.001 152 37 | 0.022 749 32 | 0.020 307 7  |
| KEGG    | hsa05142   | 美洲锥虫病                  | 7/131  | 0.001 294 36 | 0.022 749 32 | 0.020 307 7  |
| KEGG    | hsa05133   | 百日咳                    | 6/131  | 0.001 402 47 | 0.022 749 32 | 0.020 307 7  |
| KEGG    | hsa05144   | 疟疾                     | 5/131  | 0.001 227 92 | 0.022 749 32 | 0.020 307 7  |
| KEGG    | hsa04062   | 趋化因子信号通路               | 10/131 | 0.001 088    | 0.022 749 32 | 0.020 307 7  |

注: BP 为生物学功能, CC 为细胞成分, MF 为分子功能, KEGG 为京都基因与基因组百科全书。

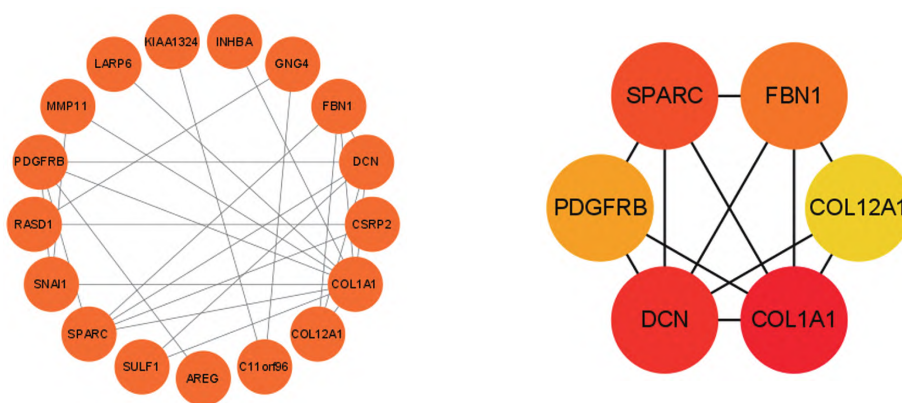


图4 关键基因构建的 PPI 网络

注: A, PPI 可视化网络图; B, 6 个关键基因。

腺癌患者预后的关系, 结果显示 COL1A1 和 COL12A1 基因的表达量升高可显著降低结直肠癌患者的 DFS 期, 而 SPARC、PDGFRB、FBN1 和 DCN 的表达水平与预后之间的关系无统计学意义 (图 5)。

### 3 讨论

在分子层面筛选出能够判断结直肠癌恶变风险的生物学标志物意义重大, 因为这将结直肠癌患者提供更合理的诊疗方案。本项研究旨

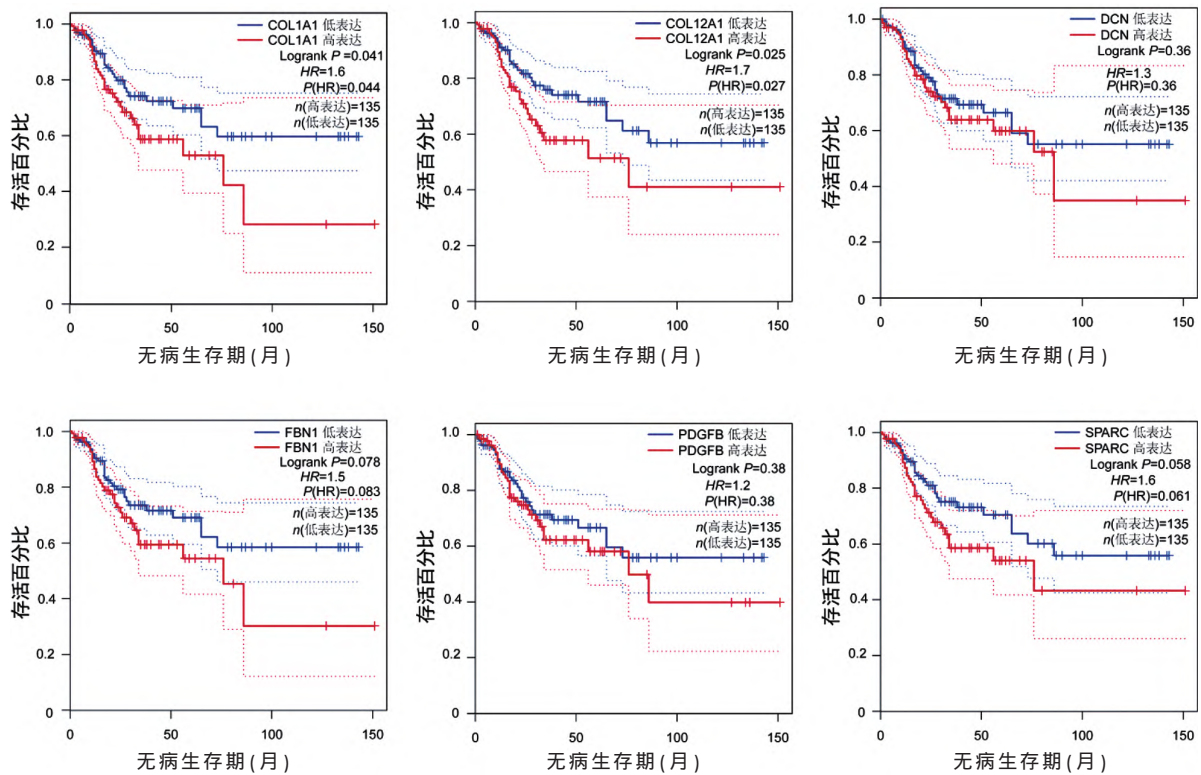


图5 主效基因的预后特征

在通过多种生物信息学方法分析并筛选结直肠腺瘤-腺癌序列中的主效基因,以发现能够帮助判断结直肠腺瘤恶变风险的生物标志物。本研究发现共有6个主效基因参与了该序列,有些基因已经被其他研究所证实在结直肠腺瘤中高表达,例如COL1A1、SPARC<sup>[8-11]</sup>,这也是对我们研究的佐证。

我们将结直肠腺瘤与腺癌对比所得到的关键差异基因行GO功能分析发现,细胞外基质组织和细胞外结构组织是两个最重要的生物学过程,细胞成分主要是含胶原的细胞外基质和内质网腔,细胞外基质结构成分、内肽酶活性、受体配体活性是主要的分子功能。此外,细胞外基质受体相互作用、蛋白质消化和吸收、TGF- $\beta$ 信号通路是KEGG富集的主要信号通路。我们对比分析发现,参与上述生物学过程和信号通路的基因在结直肠腺瘤中的表达量均高于腺癌,表明了细胞外基质合成的过程参与了结直肠腺瘤-腺癌序列的发展。

细胞外基质是一种上层结构,具有支撑细胞作用。另外,它还可以向细胞传递信号,调控细胞的行为。此外,细胞外基质还能够直接参与细胞恶变过程,并在恶性肿瘤发生发展过程中发挥重要作用<sup>[12]</sup>。我们分析发现,在6个主效基因中,DCN、

SPARC、COL1A1和COL12A1这4个主效基因可直接参与细胞外基质的合成,并且都在结直肠腺瘤组织中高表达。COL1A1和COL12A1已被证实在多种恶性肿瘤中表达量明显高于正常组织,并且与肿瘤细胞恶性表型及患者预后相关<sup>[13-16]</sup>。另有研究发现,COL1A1在早期结直肠癌组织中表达水平升高,且恶性肿瘤维持其侵袭性需要依赖COL1A1<sup>[17]</sup>。细胞水平的功能研究发现,COL1A1还可以诱导细胞发生上皮间充质转变样改变,并促进肿瘤进展和转移<sup>[18]</sup>。COL12A1同样也被发现在结直肠腺瘤组织中高表达,且与患者预后呈负相关<sup>[14]</sup>。本项研究同样发现COL1A1和COL12A1在结直肠腺瘤中的表达量高于腺癌,通过生存分析显示COL1A1与COL12A1表达水平升高与结直肠腺瘤患者DFS呈负相关。因此,我们认为COL1A1和COL12A1在结直肠腺瘤-腺癌序列中发挥重要作用。DCN是细胞外基质的一个结构组件,可以参与细胞增殖、迁移和分化等生物学过程<sup>[19]</sup>。虽然此前有多项研究认为DCN在结直肠癌发生发展的过程中起负性调控作用<sup>[20-21]</sup>,但也有学者发现DCN对结直肠癌则有促进作用<sup>[22]</sup>。我们的研究显示DCN在结直肠腺瘤中表达量高于腺癌,可能对结直肠

腺瘤-腺癌序列有促进作用。SPARC 是基质细胞蛋白家族的成员之一,也是一种钙结合蛋白,研究表明,SPARC 的低表达是大肠癌一个独立的不良预后因素<sup>[23]</sup>。但本研究发现,SPARC 和 DCN 的表达量与结直肠腺癌患者的无病生存期无明显相关性。因此,我们推测 COL1A1 和 COL12A1 在结直肠腺瘤-腺癌序列中发挥重要作用,并且可作为判断结直肠腺瘤恶变风险的生物学标志物。本研究不足在于虽然应用生物信息学方法初步筛选了可用于判断结直肠腺瘤恶变风险的生物标志物,但缺少实验验证,因此,本研究结果仍有待于临床大规模样本的进一步验证。

#### 参考文献

- [1] MALVEZZI M, CARIOLI G, BERTUCCIO P, et al. European cancer mortality predictions for the year 2018 with focus on colorectal cancer[J]. *Ann Oncol*, 2018, 29(4): 1016-1022.
- [2] ZHENG R, ZHANG S, ZENG H, et al. Cancer incidence and mortality in China, 2016 [J]. *JNCC*, 2022, 2(1): 1-9.
- [3] STRUM WB. Colorectal adenomas [J]. *N Engl J Med*, 2016, 374(11): 1065-1075.
- [4] SMIT WL, SPAAN CN, JOHANNES DE BOER R, et al. Driver mutations of the adenoma-carcinoma sequence govern the intestinal epithelial global translational capacity[J]. *Proc Natl Acad Sci USA*, 2020, 117(41): 25560-25570.
- [5] 罗喜顺, 王海鹏, 朱袭嘉, 等. 结直肠癌 KRA S 基因突变性与 P53 及 MCC 基因的相关性研究[J/CD]. *消化肿瘤杂志(电子版)*, 2021, 13(2): 125-129.
- [6] SHI X, GAO GY, SHEN J. Identification of microRNA Signature and Key Genes Between Adenoma and Adenocarcinomas Using Bioinformatics Analysis[J]. *Oncotargets Ther*, 2021, 14: 4707.
- [7] 国家癌症中心中国结直肠癌筛查与早诊早治指南制定专家组. 中国结直肠癌筛查与早诊早治指南(2020, 北京)[J]. *中国肿瘤*, 2021, 30(1): 1-28.
- [8] WU W, YANG Z, LONG F, et al. COL1A1 and MZB1 as the hub genes influenced the proliferation, invasion, migration and apoptosis of rectum adenocarcinoma cells by weighted correlation network analysis [J]. *Bioorg Chem*, 2020, 95: 103457.
- [9] GALAMB O, SIPOS F, SPISAK S, et al. Potential biomarkers of colorectal adenoma-dysplasia-carcinoma progression: mRNA expression profiling and in situ protein detection on TMAs reveal 15 sequentially upregulated and 2 downregulated genes [J]. *Cell Oncol*, 2009, 31(1): 19-29.
- [10] ZHANG Z, WANG Y, ZHANG J, et al. COL1A1 promotes metastasis in colorectal cancer by regulating the WNT/PCP pathway [J]. *Mol Med Rep*, 2018, 17(4): 5037-5042.
- [11] WANG Q, SHI L, SHI K, et al. CircCSPP1 functions as a ceRNA to promote colorectal carcinoma cell EMT and liver metastasis by upregulating COL1A1 [J]. *Front Oncol*, 2020, 10: 850.
- [12] TZANAKAKIS G, KAVASI RM, VOUDOURI K, et al. Role of the extracellular matrix in cancer-associated epithelial to mesenchymal transition phenomenon [J]. *Dev Dyn*, 2018, 247(3): 368-381.
- [13] WW A, ZHU Y B, FL C, et al. COL1A1 and MZB1 as the hub genes influenced the proliferation, invasion, migration and apoptosis of rectum adenocarcinoma cells by weighted correlation network analysis [J]. *Bioorg Chem*, 2020, 95: 103457.
- [14] WU Y, XU Y. Integrated bioinformatics analysis of expression and gene regulation network of COL12A1 in colorectal cancer [J]. *Cancer Med*, 2020, 9(13): 4743-4755.
- [15] JIANG X, WU M, XU X, et al. COL12A1, a novel potential prognostic factor and therapeutic target in gastric cancer [J]. *Mol Med Rep*, 2019, 20(4): 3103-3112.
- [16] MA H P, CHANG HL, BAMODU OA, et al. Collagen 1A1 (COL1A1) Is a Reliable Biomarker and Putative Therapeutic Target for Hepatocellular Carcinogenesis and Metastasis [J]. *Cancers (Basel)*, 2019, 11(6): 786.
- [17] ZOU X, FENG B, DONG T, et al. Up-regulation of type I collagen during tumorigenesis of colorectal cancer revealed by quantitative proteomic analysis [J]. *J Proteomics*, 2013, 94(Complete): 473-485.
- [18] LIU J, SHEN JX, WU HT, et al. Collagen 1A1 (COL1A1) promotes metastasis of breast cancer and is a potential therapeutic target [J]. *Discov Med*, 2018, 25(139): 211-223.
- [19] JÄRVINEN TA, PRINCE S. Decorin: a growth factor antagonist for tumor growth inhibition [J]. *Biomed Res Int*, 2015, 2015: 654765.
- [20] LIU Z, YANG Y, ZHANG X, et al. An oncolytic adenovirus encoding decorin and granulocyte macrophage colony stimulating factor inhibits tumor growth in a colorectal tumor model by targeting pro-tumorigenic signals and via immune activation [J]. *Hum Gene Ther*, 2017, 28(8): 667-680.
- [21] RESZEGI A, HORVÁTH Z, KARÁSZI K, et al. The protective role of decorin in hepatic metastasis of colorectal carcinoma [J]. *Biomolecules*, 2020, 10(8): 1199.
- [22] ŽLAJPAH M, URH K, GROSEK J, et al. Differential Expression of Decorin in Metastasising Colorectal Carcinoma Is Regulated by miR-200c and Long Non-Coding RNAs [J]. *Biomolecules*, 2022, 10(1): 142.
- [23] DREV D, HARPAIN F, BEER A, et al. Impact of Fibroblast-Derived SPARC on Invasiveness of Colorectal Cancer Cells. *Cancers (Basel)*, 2019, 11(10): 1421.