DOI:10.16016/j.2097-0927.202201028

综合生物信息学分析鉴定与胃癌预后相关关键基因

唐永曜[1],宋 晶[1],缪诗琪[2],蔡 璟[1],宋方洲[1] 400016 重庆,重庆医科大学基础医学院:分子医学与肿瘤研究中心[1],生物信息教研室[2]

[**摘要**] **目的** 综合生物信息学方法分析识别与胃癌预后相关关键基因,探讨其能否作为胃癌预后预测的潜在生物标志物。**方法** 从美国高通量基因表达(Gene Expression Omnibus, GEO)数据库下载胃癌患者基因芯片数据集(GSE79973)并在癌症和肿瘤基因图谱(The Cancer Genome Atlas, TCGA)数据库中获取胃癌转录组数据和相应临床资料。利用 R 软件中的 Limma 和 WGCNA 程序包筛选差异共表达基因。使用 ClusterProfler 程序包对差异共表达基因进行 GO 和 KEGG 富集分析。采用 STRING 数据库对差异共表达基因进行蛋白质相互作用(protein-protein interaction, PPI)网络分析,并用 Cytoscape 软件筛选关键基因。通过在线工具 GEPIA2 验证关键基因的转录表达情况,利用 Kaplan-Meier 单因素生存分析鉴定与胃癌预后相关的关键基因。最后采用 RT-qPCR 和免疫组化方法验证预后相关关键基因的差异表达。**结果** 共筛选出 18 个差异共表达基因,通过 PPI 网络分析,识别出 10 个关键基因(*TNFRSF*11*B*、*HOXC*10、*HOXA*10、*TEAD*4、*GKN*2、*GKN*1、*SALL*4、*SGK*1、*ANGPT*2、*NKX*6-2)。生存分析结果显示 *HOXA*10($P=0.014$)和 *ANGPT*2($P=0.002$)的高表达可能与胃癌患者预后生存较差有关,RT-qPCR 和免疫组化结果显示其在胃癌组织中高表达($P<0.05$)。**结论** *HOXA*10 和 *ANGPT*2 是胃癌预后相关关键基因,可作为胃癌临床治疗和预后预测的潜在生物标志物。

[**关键词**] 生物信息学;胃癌;预后;加权基因共表达网络分析;差异共表达基因
[**中图法分类号**] R318.04;R730.7;R735.2 [**文献标志码**] A

Identification of hub genes associated with prognosis of gastric cancer by integrated bioinformatics analysis

TANG Yongyao[1], SONG Jing[1], MIAO Shiqi[2], CAI Jing[1], SONG Fangzhou[1] [1]Molecular and Tumor Research Center, [2]Department of Bioinformatics, College of Basic Medical Sciences, Chongqing Medical University, Chongqing, 400016, China

[**Abstract**] **Objective** To identify the hub genes associated with development of gastric cancer (GC) by integrated bioinformatics analysis, and investigate whether these genes can be used as latent biomarkers of predicting GC prognosis. **Methods** The gastric cancer of microarray data was downloaded from the Gene Expression Omnibus (GEO) database (GSE79973), and the transcriptome data and corresponding clinical data of GC patients were obtained from The Cancer Genome Atlas (TCGA) database. Differential gene expression analysis and Weighted Gene Co-expression Network Analysis (WGCNA) were performed on the dataset using the Limma and WGCNA packages in the R software, respectively, in order to screen differentially co-expressed genes. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis of differential co-expression genes were performed using the ClusterProfler package. The search tool for the retrieval of interacting genes (STRING) was carried out the protein-protein interaction (PPI) network of differential co-expression genes, and the software Cytoscape was used to visualize

[**基金项目**] 重庆市技术创新与应用发展专项重点项目(cstc2018jscx-mszdX0031)
[**通信作者**] 宋方洲,E-mail:fzsong@cqmu.edu.cn

2097-0927/Copyright © 2022, 陆军军医大学(第三军医大学)。这篇开放获取文章遵循 CC BY 许可协议(https://creativecommons.org/licenses/by/4.0/)

and screen out hub genes. The transcriptional expression of the hub gene was further verified by GEPIA2 database. Survival analysis was performed by using the survival package in R software to identify prognostic hub genes. Finally, RT-qPCR and immunohistochemical assay were used to verify the differential expression of hub genes related to prognosis. **Results** A total of 18 differential co-expression genes were screened out in this study, and PPI network analysis identified 10 key genes (*TNFRSF11B*, *HOXC10*, *HOXA10*, *TEAD4*, *GKN2*, *GKN1*, *SALL4*, *SGK1*, *ANGPT2* and *NKX6-2*). The results of survival analysis found that the high expression of *HOXA10* ($P=0.014$) and *ANGPT2* ($P=0.002$) may be related to the survival and prognosis of GC patients, and RT-qPCR and immunohistochemical assay further verified the high expression of these 2 molecules in GC tissues ($P<0.05$). **Conclusion** *HOXA10* and *ANGPT2* are hub genes for prognosis of GC, and may be used as latent biomarkers of prediction.

[**Key words**] bioinformatics; gastric cancer; prognosis; weighted gene co-expression network analysis; differentially co-expressed genes

Supported by the Key Project of Technology Innovation and Application Development Program of Chongqing (cstc2018jscx-mszdX0031). Corresponding author: SONG Fangzhou, E-mail: fzsong@cqmu.edu.cn

胃癌是全世界常见的恶性肿瘤,其发病率位居恶性肿瘤的第5位,且每年全球新增的胃癌病例中超过40%来自中国[1-2]。胃癌在早期通常无任何症状,部分患者确诊时已处于晚期,其生存率仅为24%。近年来开发出针对中晚期胃癌的一些靶向药物并批准用于临床,包括曲妥珠单抗(HER2阳性患者第一线)、雷莫西单抗(抗血管生成第二线)和nivolumab或pembrolizumab(抗PD-1第三线)药物[3-4]。然而,胃癌患者总体预后仍然较差。因此,开发有效的生物标记物用于胃癌预后预测及作为靶向治疗的靶标十分重要。

由于基因组技术的进步,通过生物信息学对基因表达谱的分析在分子机制研究中越来越受欢迎,并且在发现疾病特异性生物标志物方面也发挥着越来越重要的作用[5]。从全基因组表达谱中了解基因功能和基因关联的一个重要方法就是加权基因共表达网络分析(weighted gene co-expression network analysis, WGCNA)[6]。WGCNA可用于检测高度相关基因的共表达模块和与感兴趣的临床性状相关模块[7],为预测共表达基因的功能和寻找在人类疾病中起关键作用的基因提供了新的途径[8-9]。此外,转录组学中另一个强有力的分析方法就是差异基因表达分析。它为研究基因组如何调控生物表型的分子机制以及发现病变组织和正常组织基因表达水平的定量变化提供了方法[10]。基因的不同表达有助于为某些疾病寻找潜在的生物标志物。因此,将WGCNA和差异基因表达分析的相结合,寻找差异共表达基因,为筛选肿瘤生物标志物提供新的方法。

本研究从癌症和肿瘤基因图谱(The Cancer Genome Atlas, TCGA)和高通量基因表达(Gene Expression Omnibus, GEO)数据库中获取胃癌患者的mRNA数据,采用WGCNA联合差异基因表达分析方法,获得差异共表达基因。并对差异共表达基因进行一系列生物信息学分析,挖掘出与胃癌预后相关的关键基因;最后在胃癌组织样本中验证其表达,以期寻找胃癌预后预测的潜在生物标志物。

1 材料与方法

1.1 数据及样本获取

胃癌患者转录组数据和对应临床数据(TCGA-STAD)通过R软件TCGAbiolinks程序包从TCGA数据库(https://portal.gdc.cancer.gov/)中下载[11],包括375例胃癌样本和32例正常样本。此外,还从GEO数据库(http://www.ncbi.nlm.nih.gov/geo/)下载胃癌患者芯片表达谱GSE79973,其中包含10例胃癌样本和10例正常样本,芯片平台为Affymetrix Human Genome U133 Plus 2.0 Array[12]。根据制造商提供的注释文件将探针转换为基因名称,同一基因的重复探针被移除。保留了所有重复探针的中位数表达式值。最终,选择24 654个基因的列表进行后续分析。胃癌组织cDNA芯片购自上海芯超生物科技有限公司,其中胃癌组织cDNA芯片包含30例胃癌组织(含配对的癌旁组织)及临床病理预后信息。

收集2020-2021年重庆医科大学第一附属医院94例胃癌患者的样本(其中含配对的癌旁组织86例)。患者术前均未接受过放疗或化疗等。本研究方案经重庆医科大学伦理委员会批准(2020008),遵循赫尔辛基宣言的原则,所有参与者已签署书面知情同意书。

1.2 差异表达分析

在 R 软件中采用 Limma 程序包标准化矩阵数据[13],鉴定每个数据集的肿瘤组织和癌旁组织之间的差异基因(differentially expressed genes,DEGs)。使用 Bonferroni-Holm (BH)法调整 P 值。调整后 $P<0.05$ 和 $|\log_2 \text{foldchange}|>1$ 表示基因差异具有统计学意义。通过使用 R 软件中的 ggplot2 包[14],将 TCGA-STAD 和 GSE79973 数据集的 DEGs 可视化为火山图和热图。

1.3 加权基因共表达网络分析及差异共表达基因的获取

共表达网络促进了基于网络的基因筛选方法,可用于识别候选生物标志物和治疗靶点。将 TCGA-STAD 和 GSE79973 的基因表达谱分别在 R 软件中利用 WGCNA 程序包构建共表达网络[6]。首先计算各个基因之间的 Pearson 系数使其转化为相似矩阵,通过 WGCNA 包的 pick soft threshold 函数自动进行网络拓扑分析选择软阈值 β,β 可以强调基因之间的强弱相关性。确定 β 后相似矩阵转化为邻接矩阵,再将邻接矩阵转化为拓扑重叠矩阵(topological overlap matrix,TOM),设置模块最小基因数为 50,剪切高度为 0.25,通过层次聚类是表达相近的基因置于同一基因模块,并将阈值设置为 20 000 以消除异常值,利用动态混合切割方法,将表达模式类似的基因分到不同模块中。计算基因模块和表型(癌组织和正常样本)的 Pearson 相关系数,选择与临床性状高度相关的模块作为候选模块[15]。随后,将候选模块中的共表达基因与之前的差异基因取利用 R 软件的 VennDiagram 程序包构建韦恩图取交集,得到差异共表达基因,作为潜在预后相关基因进行下一步分析[16]。

1.4 差异共表达基因的功能富集分析

R 软件中的 ClusterProfler 程序包对差异共表达基因进行 GO 富集分析,包括生物过程、细胞成分和分子功能[17-18]。同样再对差异共表达基因进行 KEGG 富集分析[19]。

1.5 蛋白质相互作用(protein-protein interaction,PPI)网络的构建与关键基因的识别

使用在线数据库 STRING(Search Tool for the Retrieval of Interacting Genes)构建差异共表达基因之间的 PPI 网络,置信评分≥0.4[20]。随后通过软件 Cytoscape(v3.82)可视化网络[21]。在共表达网络中,MCC 算法算是查找关键节点最有效的方法[22]。每个节点的 MCC 由 Cytoscape 中的插件 cytohubba 计算并将前十名的视作关键基因。

1.6 关键基因表达水平及预后价值的验证

为了验证关键基因的可靠性,使用在线工具 GEPIA2 验证 TCGA-STAD 数据中肿瘤组织与正常组织关键基因的 mRNA 表达水平[23]。然后,基于 TCGA-STAD 数据,利用 R 软件中的生存包进行 Kaplan-Meier 单因素生存分析,探索患者的整体生存(overall survival, OS)与关键基因之间的关系。本研究仅选择完成随访时间的患者进行生存分析,然后根据关键基因的中位表达值分为 2 组。$P<0.05$ 的预后相关关键基因被认为具有统计学意义。

1.7 采用 RT-qPCR 在胃癌患者 cDNA 芯片中验证预后相关关键基因的表达

1.7.1 RNA 抽提 ①将待研磨的组织样品从液氮或者-80 ℃冰箱中取出,无菌刀片于干冰上将组织样品切割成约 3 mm×3 mm×3 mm 大小,置于装有 1 mL TRIzol 裂解液的 1.5 mL EP 管中。将超细匀浆机工作头浸入 TRIzol 裂解液中 5~10 s 灭活 RNA 酶后,研磨组织;研磨后 4 ℃、5 000 r/min 离心 3 min,弃沉淀,吸取上清至新的 1.5 mL EP 管中,室温放置 10 min。②每 1 000 μL TRIzol 加入 200 μL 1-溴-3-氯丙烷,剧烈震荡 EP 管 15 s,室温静置 10 min。③4 ℃、12 000 r/min 离心 15 min。④吸取上层上清至新的 1.5 mL EP 管,加入 1.5 倍体积预冷的异丙醇,混匀后 4 ℃静置 20 min。⑤4 ℃、12 000 r/min 离心 15 min,弃上清。⑥加入 500 μL 75%乙醇,洗涤沉淀。⑦4 ℃、12 000 r/min 离心 5 min,弃去大部分上清。⑧4 ℃、12 000 r/min 再次离心 1 min,弃去上清,室温干燥。⑨待 RNA 沉淀基本透明时,加入 RNase-free 水(加入体积视 RNA 沉淀量而定)至完全溶解。⑩ RNA 纯化:在上述溶解的 RNA 溶液中加入 2.5 μL DNase 和 10 μL RDD buffer,并加水补足到 100 μL,再次重复上述①~⑨步骤,Nanodrop one 分析测定所抽提 RNA 的浓度及质量,安捷伦 2100 生物分析仪检测 RNA 完整性,RNA 完整值(RNA integrity number, RIN)大于 7 的样本可进行后续实验。

1.7.2 RNA 逆转录获得 cDNA 根据 TaKaRa 公司 PrimeScript™ RT Master Mix (Perfect Real Time) 的操作指南,具体步骤如下:①将 2 μL 5× PrimeScript RT Master Mix(Perfect Real Time)和 500 ng 总 RNA 加入到 PCR 管中,补充 RNase-Free H_2O 至 10 μL,轻柔混匀后离心。②按照 37 ℃ 15 min(逆转录反应),85 ℃ 5 s(逆转录酶的失活反应),4 ℃程序进行逆转录。③得到的 RT 产物 cDNA 可用于制备 cDNA 芯片。

1.7.3 RT-qPCR 反应 以 cDNA 芯片为模板,根据诺唯赞公司 ChamQ Universal SYBR qPCR Master Mix 试剂盒说明书进行操作。以 β-actin 作为内参,数据计算采用 $2^{-\Delta\Delta Ct}$ 方法。

1.8　免疫组织化学在胃癌患者组织切片中验证预后相关关键基因的表达

①将胃癌石蜡切片脱蜡至水。②3%双氧水室温孵化5 min,消除内源性过氧化物酶的活性。③蒸馏水冲洗,PBS浸泡5 min,重复2次。④10%山羊血清(用PBS稀释)封闭,室温孵育10 min,小心倒掉血清,加入一抗工作液,4 ℃孵化过夜。⑤PBS冲洗,5 min,重复3次。⑥滴加1 mL生物素标记二抗工作液,37 ℃孵育20 min。⑦PBS冲洗5 min,重复3次。⑧滴加1 mL辣根酶标记的链霉卵白素工作液,37 ℃孵育20 min。⑨PBS冲洗,5 min,重复3次。⑩DAB显色剂显色,每30秒后立即显微镜观察有无染色,至2 min时变为灰棕色提示着色。⑪自来水充分冲洗,复染,脱水,透明,封片。免疫组化评分标准:染色强度得分×阳性比例得分。

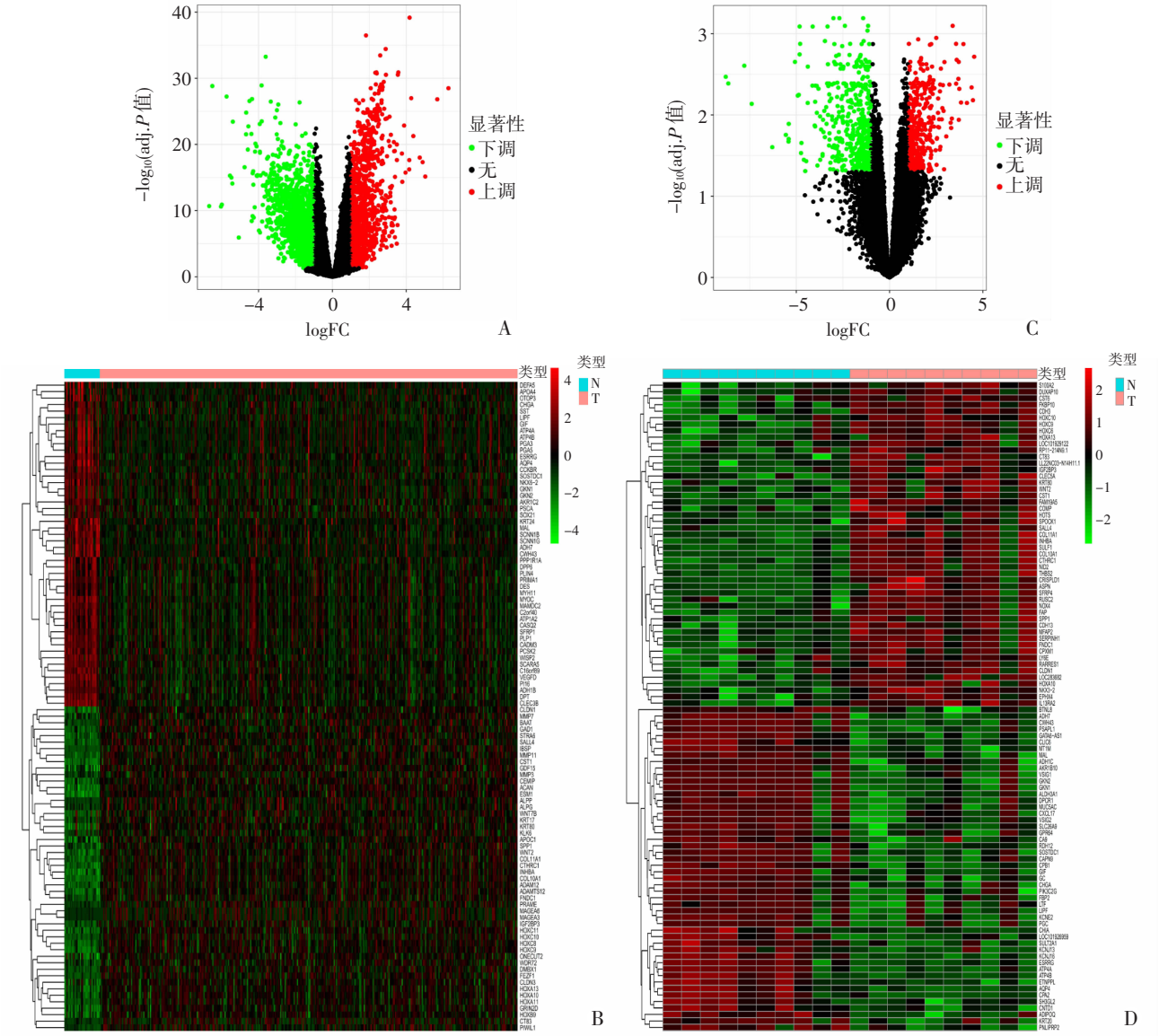
1.9　统计学分析

采用R软件4.0.2版本(https://www.r-project.org/)处理实验数据。计量资料以$\bar{x}\pm s$表示,独立样本采用独立样本t检验,配对样本采用配对样本t检验。Kaplan-Meier生存分析采用Log-rank检验。$P<0.05$表示差异具有统计学意义。

2　结果

2.1　差异基因的鉴定

在TCGA数据集中一共识别出2 694个DEGs,包括1 124个上调基因和1 570个下调基因(图1A、B)。进一步在GEO数据集GSE79973中识别出857个DEGs,其中410个表达量上调,447个表达量下调(图1C、D)。



A: TCGA-STAD差异基因火山图;B: TCGA-STAD差异基因表达热图;C: GSE79973差异基因火山图;D: GSE79973差异基因表达热图

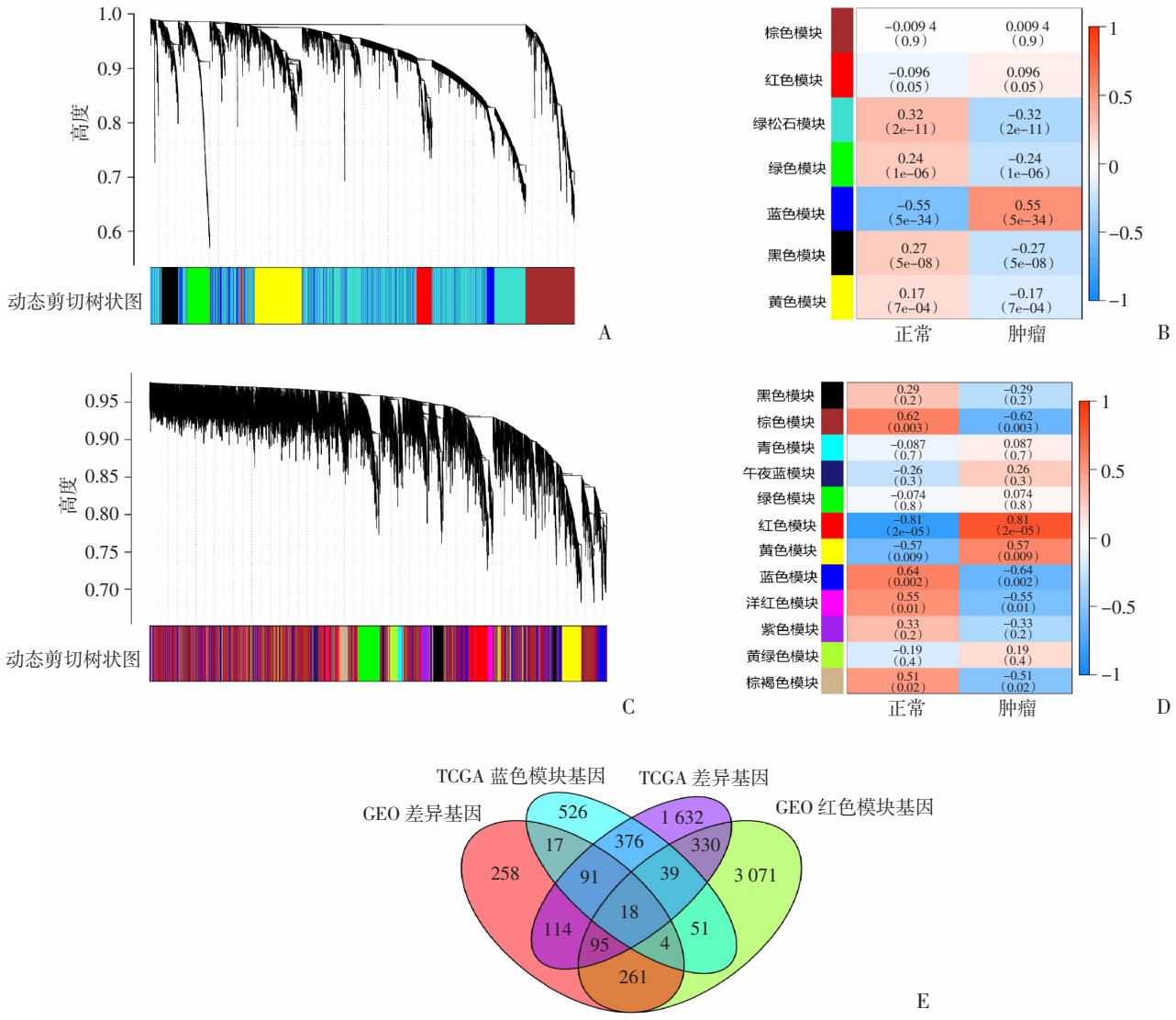
图1　差异表达基因的鉴定

2.2 加权基因共表达模块的构建及差异共表达基因的获取

为了研究胃癌患者的功能聚类,利用 WGCNA 软件包从 TCGA-STAD 和 GES79973 数据集构建共表达网络。结果在 TCGA-STAD 中确定 7 个模块(图 2A)以及在数据集 GSE79973 中确定 12 个模块(图 2C),绘制模块-性状的关系见图 2B、D,以评估每个模块与两个临床性状(癌症和正常)之间的关系。TCGA-STAD 中的蓝色模块($r=0.55$, $P=5e-34$)和 GSE79973 的红色模块($r=0.81$, $P=2e-05$)和胃癌相关最高。分别在 TCGA 数据集的蓝色模块和 GSE79973 的红色模块中发现了 1 176 个和 3 869 个共表达基因,并与之前的 DEGs 取交集,识别出 18 个差异共表达基因(图 2E)。

2.3 差异共表达基因的功能富集分析

为了进一步了解差异共表达基因的潜在功能,采用 clusterProfiler 包对其进行了基因富集分析。GO 富集分析显示,18 个差异共表达基因在生物学过程(BP)中主要富集于肢体发育和胚胎附肢形态发生,在细胞成分(CC)的结果中,揭示了这些基因主要参与转录调节复合物。此外,在分子功能(MF)分析中,氧化还原酶活性和氧化金属离子被认为与 18 个基因有关(图 3A)。在 KEGG 通路富集分析中,这 18 个基因主要富集在细胞周期通路中(图 3B)。

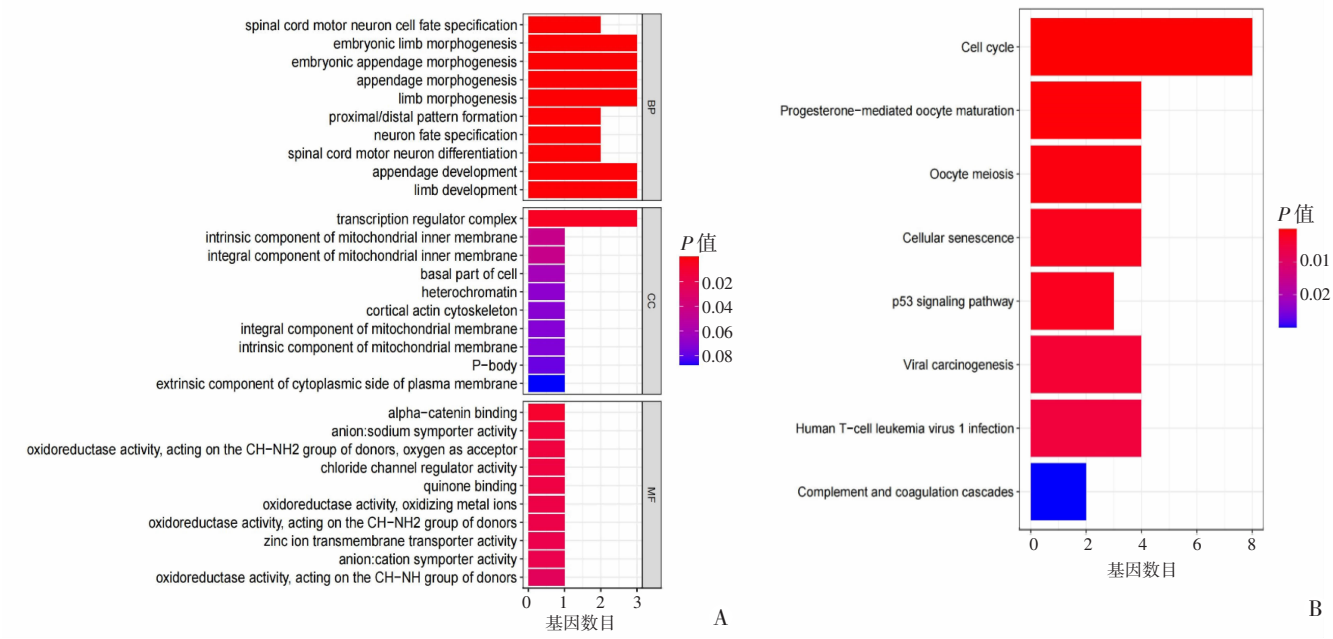


A:TCGA-STAD 数据集模块聚类树状图;B:TCGA-STAD 数据集基因模块和临床性状相关性 括号内数据为对应相关系数的 P 值;C:GSE79973 模块聚类树状图;D:GSE79973 基因模块和临床性状相关性 括号内数据为对应相关系数的 P 值;E:共表达基因和差异基因交集的韦恩图

图 2 差异共表达基因的识别

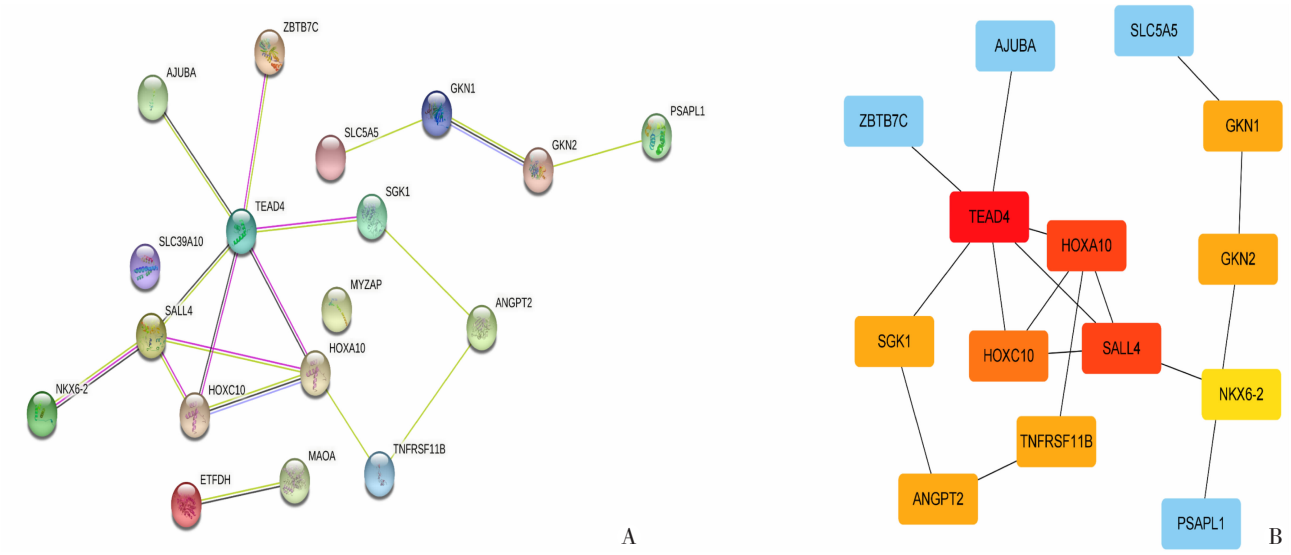
2.4 PPI 网络的构建和关键基因的鉴定

利用 STRING 数据库建立了差异共表达基因之间的 PPI 网络,具有 18 个节点和 17 个边缘(图 4A)。通过利用 CytoHubba 插件的 MCC 算法从 PPI 网络中选取的关键枢纽基因(图 4B)。根据 MCC 评分,得分排名前十的 *NKX6-2*、*GKN2*、*GKN1*、*SGK1*、*TNFRSF11B*、*ANGPT2*、*HOXC10*、*SALL4*、*HOXA10*、*TEAD4* 基因被选作关键基因。



A:GO 富集分析;B:KEGG 富集分析

图 3 差异共表达基因的功能富集分析

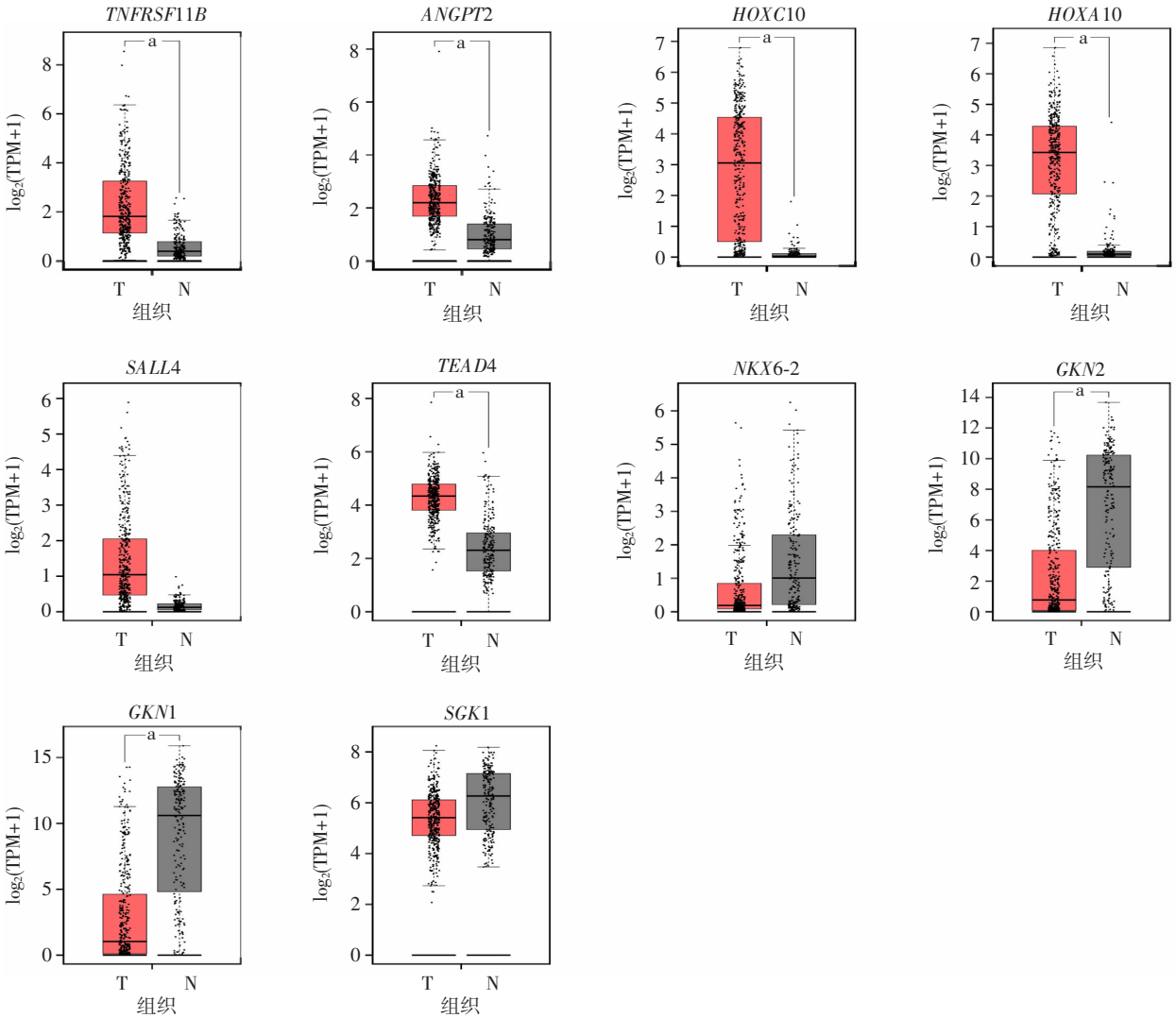


A: 通过数据库 STRING 构建的 18 个重叠基因的 PPI 网络包括 18 个节点和 17 个边缘 PPI 网络;B:使用来自 PPI 网络的 MCC(最大集团中心性)算法鉴定关键基因 在 MCC 中,得分高的基因由红色节点表示,得分低的基因由黄色节点表示

图 4 PPI 网络可视化和潜在关键基因

2.5 关键基因表达水平、预后价值的验证

在 GEPIA2 工具内验证 10 个关键基因(*NKX*6-2、*GKN*2、*GKN*1、*SGK*1、*TNFRSF*11*B*、*ANGPT*2、*HOXC*10、*SALL*4、*HOXA*10 和 *TEAD*4)的 mRNA 表达水平。如图 5 所示,*TNFRSF*11*B*、*ANGPT*2、*HOXC*10、*HOXA*10、*SALL*4 和 *TEAD*4 在胃癌中显著上调,而 *NKX*6-2、*GKN*2、*GKN*1、*SGK*1 在胃癌中下调。此外,使用 R 软件中生存包(图 6)对 10 个关键基因进行了 OS 分析,以研究胃癌患者中关键基因的预后价值。根据中位基因表达值将患者分为高表达组和低表达组。Kaplan-Meier 分析表明,在 10 个关键基因中,*ANGPT*2 和 *HOXA*10 的较高表达水平与胃癌患者的 OS 较差显著相关(Log-rank $P<0.05$,图 6)。RT-qPCR 在 cDNA 芯片中检测上述 2 个基因的相对表达量,比较癌和癌旁的表达差异(图 7)。在 30 例胃癌 cDNA 芯片中,27 例都呈现 *ANGPT*2 高表达(图 7A),20 例都呈现 *HOXA*10 高表达(图 7B)。对胃癌组织进行免疫组织化学染色,结果显示 *ANGPT*2 和 *HOXA*10 在胃癌组织的免疫组化评分均高于癌旁组织(图 8)。



T:胃癌组织($n=408$);N:正常组织($n=211$);a:$P<0.05$

图 5 从 TCGA 数据库中验证来自胃癌患者组织和正常组织中 10 个关键基因的表达水平

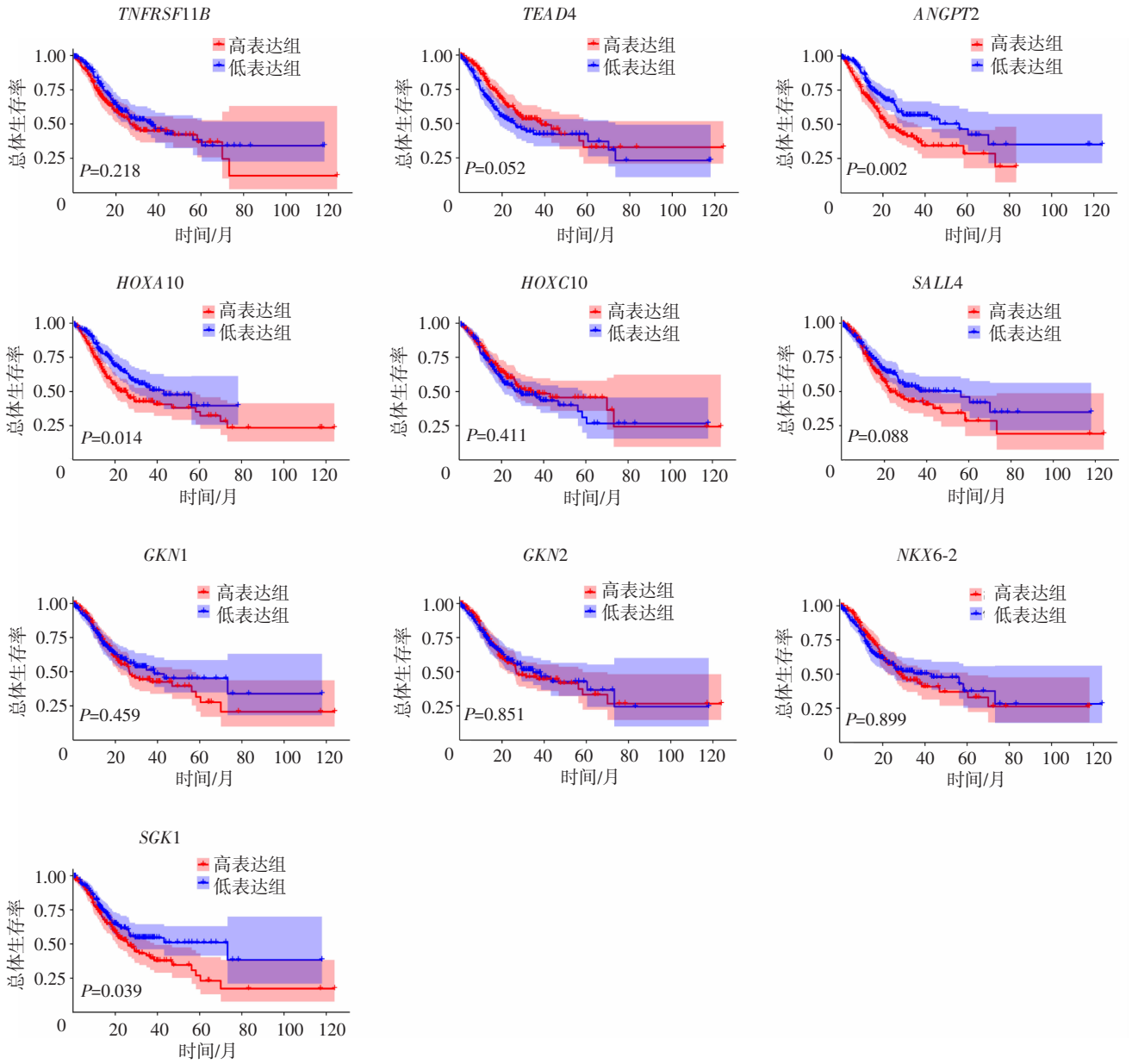
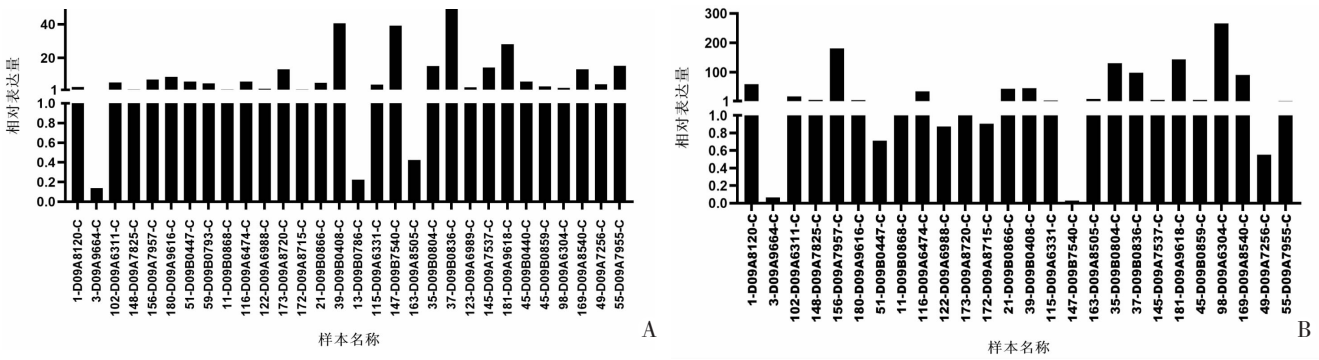
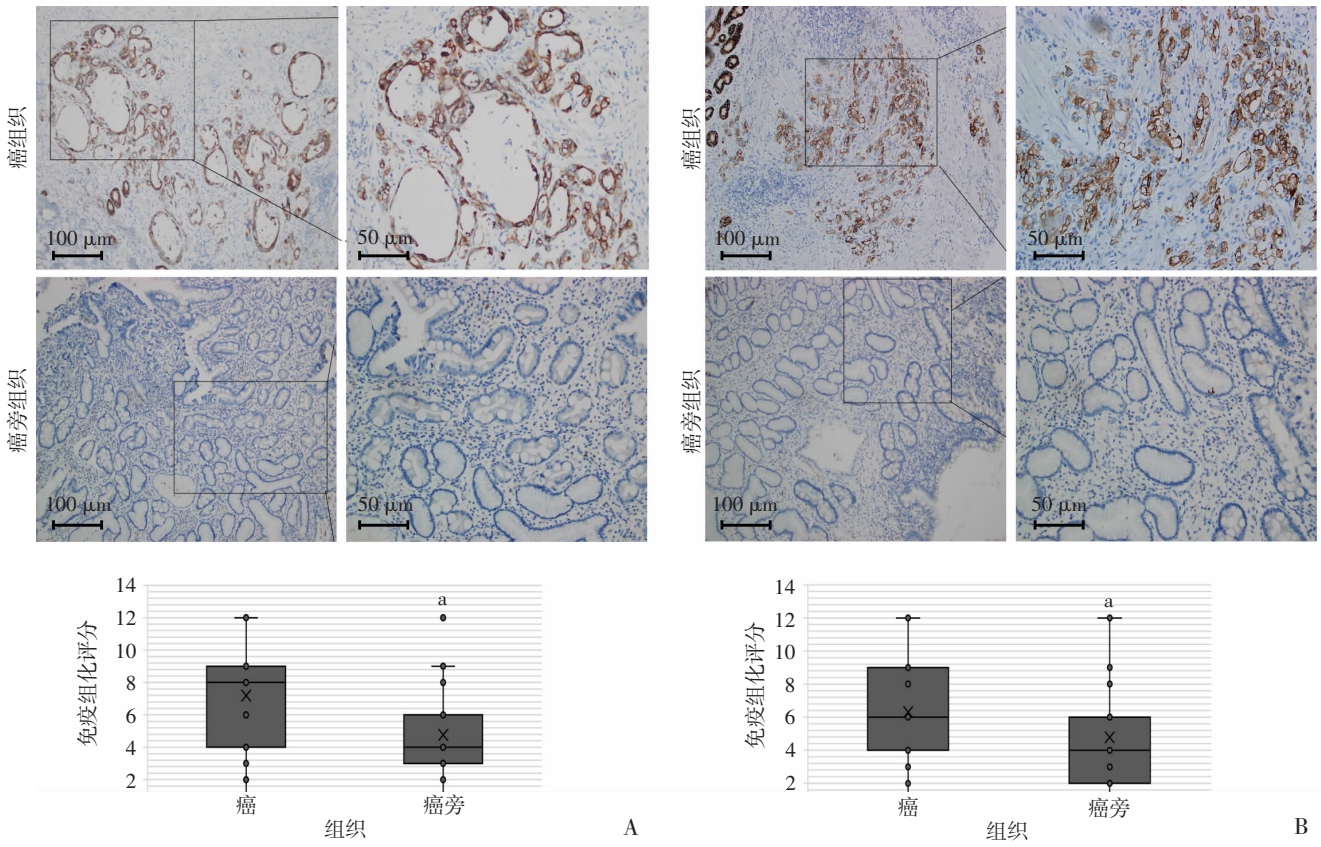


图 6　胃癌患者预后相关的 10 个关键基因生存分析



A:*ANGPT2*;B:*HOXA10*

图 7　RT-qPCR 验证胃癌患者预后相关关键基因表达



a:$P<0.05$,与癌组织比较

A:*ANGPT2* 在胃癌和癌旁组织中的表达及评分情况;B:*HOXA10* 在胃癌和癌旁组织中的表达及评分情况

图 8 免疫组化观察胃癌患者预后相关关键基因在癌组织和癌旁组织中的表达

3 讨论

胃癌是全世界高发的恶性肿瘤,被发现时多处于晚期,进展期患者预后较差,且对化疗不敏感。因此,鉴定有效的生物标志物来预测胃癌的预后及作为潜在靶标非常重要。本研究通过全面的生物信息学分析,采用 WGCNA 与差异基因表达分析相结合的方式,利用 TCGA-STAD 和 GSE79973 数据库中的胃癌患者数据鉴定出 18 个重要基因。这些基因主要富集在细胞周期等重要途径中。此外,根据 Cytoscape 中 cytohubba 插件的 MCC 评分,筛选了前 10 个胃癌相关关键基因(*NKX6-2*、*GKN2*、*GKN1*、*SGK1*、*TNFRSF11B*、*ANGPT2*、*HOXC10*、*SALL4*、*HOXA10*、*TEAD4*)。其中,*ANGPT2* 和 *HOXA10* 的上调与胃癌总生存率差有显著相关性。最后,在胃癌组织中对 *ANGPT2* 和 *HOXA10* 进行了进一步的验证。

血管生成素-2(angiopoietin 2, ANGPT2)是一种生长因子,属于血管生成素/领带(酪氨酸激酶与 Ig 和 EGF 同源结构域)信号通路[24]。它是参与血管生成的主要途径之一。相关研究表明其过表达会起到诱导肿瘤血管生成的作用,会加快肿瘤的迁徙和转移[25]。对肿瘤血管生成和转移的研究已经证实 *ANGPT2* 在癌症中的重要性,为此,*ANGPT2* 也被提议作为不同癌症类型的生物标志物。相关研究表明,原则上,*ANGPT2* 表达水平与小细胞肺癌和非小细胞肺癌的分期成正比[26]。然而,需要进一步明确其在肺癌每种类型和亚型中的确切作用和表达水平。*ANGPT2* 在胃癌中的作用机制尚不清楚,需要更进一步的研究。HOXA 家族基因是参与细胞增殖和凋亡的关键转录因子[27]。据报道,*HOXA10* 在多种癌症中表达失调,但其在胃癌中的作用仍存在争议。例如,SENTANI 等[28]报道 *HOXA10* 在约 30% 的胃癌病例中呈阳性表达,在除肠化生外的非肿瘤性胃黏膜中很少表达,其表达与浸润深度呈显著负相关,预示具有肠黏蛋白表型的患者预后较好。尽管如此,LIM 等[29]研究表明,*HOXA10* 的高表达预示着胃癌患者的不良预后。因此,应进一步探讨 *HOXA10* 在胃癌中的作用。

本研究通过全面的生物信息学分析识别出 2 个与

胃癌预后相关的关键基因(*ANGPT2*、*HOXA10*)。有研究表明,在其他肿瘤组织中较高水平的 *ANGPT2* 或 *HOXA10* 与癌症患者的预后不良有关,这与本研究分析验证结果一致。然而,本研究存在一定局限性,尚需通过一系列实验生物学方法,进一步验证影响胃癌患者预后的生存相关基因分子机制和生物学功能。

综上所述,本研究综合生物信息分析,通过 WGCNA 与差异基因表达分析相结合方式,识别验证了与胃癌预后显著相关的关键基因 *ANGPT2* 和 *HOXA10*,为胃癌患者潜在的个体化治疗提供了依据。

参考文献:

[1] BRAY F, FERLAY J, SOERJOMATARAM I, *et al*. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA Cancer J Clin, 2018, 68(6): 394-424. DOI: 10.3322/caac.21492.

[2] ZHANG G H, LI J M, LI S J, *et al*. Exploring spatial trends and influencing factors for gastric cancer based on Bayesian statistics: a case study of Shanxi, China[J]. Int J Environ Res Public Heal, 2018, 15(9): 1824. DOI: 10.3390/ijerph15091824.

[3] BANKS M, GRAHAM D, JANSEN M, *et al*. British Society of Gastroenterology guidelines on the diagnosis and management of patients at risk of gastric adenocarcinoma[J]. Gut, 2019, 68(9): 1545-1575. DOI: 10.1136/gutjnl-2018-318126.

[4] JOSHI S S, MARON S B, CATENACCI D V. Pembrolizumab for treatment of advanced gastric and gastroesophageal junction adenocarcinoma[J]. Future Oncol, 2018, 14(5): 417-430. DOI: 10.2217/fon-2017-0436.

[5] KEERTHIKUMAR S. An introduction to proteome bioinformatics[J]. Methods Mol Biol Clifton N J, 2017, 1549: 1-3. DOI: 10.1007/978-1-4939-6740-7_1.

[6] LANGFELDER P, HORVATH S. WGCNA: an R package for weighted correlation network analysis[J]. BMC Bioinformatics, 2008, 9: 559. DOI: 10.1186/1471-2105-9-559.

[7] TIAN Z L, HE W X, TANG J N, *et al*. Identification of important modules and biomarkers in breast cancer based on WGCNA[J]. Onco Targets Ther, 2020, 13: 6805-6817. DOI: 10.2147/OTT.S258439.

[8] LI J Q, ZHOU D D, QIU W L, *et al*. Application of weighted gene co-expression network analysis for data from paired design[J]. Sci Rep, 2018, 8(1): 622. DOI: 10.1038/s41598-017-18705-z.

[9] YANG Y, HAN L, YUAN Y, *et al*. Gene co-expression network analysis reveals common system-level properties of prognostic genes across cancer types[J]. Nat Commun, 2014, 5: 3231. DOI: 10.1038/ncomms4231.

[10] SEGUNDO-VAL I S, SANZ-LOZANO C S. Introduction to the gene expression analysis[J]. Methods Mol Biol, 2016, 1434: 29-43. DOI: 10.1007/978-1-4939-3652-6_3.

[11] COLAPRICO A, SILVA T C, OLSEN C, *et al*. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data[J]. Nucleic Acids Res, 2016, 44(8): e71. DOI: 10.1093/nar/gkv1507.

[12] DAVIS S, MELTZER P S. GEOquery: a bridge between the gene expression omnibus (GEO) and BioConductor[J]. Bioinformatics, 2007, 23(14): 1846-1847. DOI: 10.1093/bioinformatics/btm254.

[13] 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. DOI: 10.1093/nar/gkv007.

[14] STEENWYK J L, ROKAS A. Ggpubfigs: colorblind-friendly color palettes and ggplot2 graphic system extensions for publication-quality scientific figures[J]. Microbiol Resour Announc, 2021, 10(44): e0087121. DOI: 10.1128/MRA.00871-21.

[15] WANG C C N, LI C Y, CAI J H, *et al*. Identification of prognostic candidate genes in breast cancer by integrated bioinformatic analysis[J]. J Clin Med, 2019, 8(8): 1160. DOI: 10.3390/jcm8081160.

[16] CHEN H B, BOUTROS P C. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R[J]. BMC Bioinformatics, 2011, 12: 35. DOI: 10.1186/1471-2105-12-35.

[17] CONSORTIUM T G O. The Gene Ontology Resource: 20 years and still GOing strong[J]. Nucleic Acids Res, 2019, 47(D1): D330-D338. DOI: 10.1093/nar/gky1055.

[18] WU T Z, HU E Q, XU S B, *et al*. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data[J]. Innovation (N Y), 2021, 2(3): 100141. DOI: 10.1016/j.xinn.2021.100141.

[19] KANEHISA M, FURUMICHI M, TANABE M, *et al*. KEGG: new perspectives on genomes, pathways, diseases and drugs[J]. Nucleic Acids Res, 2017, 45(D1): D353-D361. DOI: 10.1093/nar/gkw1092.

[20] SZKLARCZYK D, GABLE A L, LYON D, *et al*. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide exper-

imental datasets[J]. Nucleic Acids Res, 2019, 47(D1): D607-D613. DOI: 10.1093/nar/gky1131.

[21] DONCHEVA N T, MORRIS J H, GORODKIN J, *et al*. Cytoscape StringApp: network analysis and visualization of proteomics data[J]. J Proteome Res, 2019, 18(2): 623-632. DOI: 10.1021/acs.jproteome.8b00702.

[22] 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. DOI: 10.1186/1752-0509-8-S4-S11.

[23] TANG Z F, KANG B X, LI C W, *et al*. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis[J]. Nucleic Acids Res, 2019, 47(W1): W556-W560. DOI: 10.1093/nar/gkz430.

[24] AKWII R G, SAJIB M S, ZAHRA F T, *et al*. Role of angiopoietin-2 in vascular physiology and pathophysiology[J]. Cells, 2019, 8(5): 471. DOI: 10.3390/cells8050471.

[25] LEWIS C E, FERRARA N. Multiple effects of angiopoietin-2 blockade on tumors[J]. Cancer Cell, 2011, 19(4): 431-433. DOI: 10.1016/j.ccr.2011.03.016.

[26] OZTUTGAN T, DEMIRER E, TAS D, *et al*. A comparative analysis of angiopoietin 2 immunohistochemical staining in various stages of lung cancer[J]. Niger J Clin Pract, 2016, 19(6): 725-729. DOI: 10.4103/1119-3077.193406.

[27] GUO C, JU Q Q, ZHANG C X, *et al*. Overexpression of HOXA10 is associated with unfavorable prognosis of acute myeloid leukemia[J]. BMC Cancer, 2020, 20(1): 586. DOI: 10.1186/s12885-020-07088-6.

[28] SENTANI K, OUE N, NAITO Y, *et al*. Upregulation of HOXA10 in gastric cancer with the intestinal mucin phenotype: reduction during tumor progression and favorable prognosis[J]. Carcinogenesis, 2012, 33(5): 1081-1088. DOI: 10.1093/carcin/bgs121.

[29] LIM J Y, YOON S O, SEOL S Y, *et al*. Overexpression of miR-196b and HOXA10 characterize a poor-prognosis gastric cancer subtype[J]. World J Gastroenterol, 2013, 19(41): 7078-7088. DOI: 10.3748/wjg.v19.i41.7078.

(收稿:2022-01-05;修回:2022-02-25)

(编辑 汪勤俭)