

# 视网膜母细胞瘤发生发展的关键基因挖掘及验证

黄子珊, 符馨予, 周希媛\*

重庆医科大学附属第二医院眼科/眼科学重庆市市级重点实验室, 重庆 400010

[中图分类号] R774.1

[文献标志码] A

[DOI]

10.11855/j.issn.0577-7402.2022.01.0001

[声明]

本文所有作者声明无利益冲突

[引用本文]

黄子珊, 符馨予, 周希媛. 视网膜母细胞瘤发生发展的关键基因挖掘及验证[J]. 解放军医学杂志, 2022, 47(1): 1-11.

[收稿日期] 2021-06-22

[录用日期] 2021-09-07

[上线日期] 2021-12-01

**[摘要]** **目的** 探索与视网膜母细胞瘤(RB)发生发展相关的关键基因。**方法** 从GEO数据库中下载3个RB基因表达谱芯片数据集, 合并后用R软件sva包消除批次效应、limma包分析RB的差异表达基因(DEGs)、clusterProfiler包对DEGs进行GO富集和KEGG通路分析, 采用STRING数据库和Cytoscape软件构建蛋白相互作用(PPI)网络并利用cytoHubba寻找网络核心基因; 对加权基因共表达网络(WGCNA)进行分析, 将基因划分模块与临床信息进行关联以找到关键模块, 进一步寻找关键模块中的关键基因。体外培养Y79、WERI-RB-1和ARPE-19细胞, 采用实时荧光定量PCR(qRT-PCR)技术检测5个关键基因的mRNA表达量; 收集20例RB组织样本, 利用免疫组化技术检测关键蛋白的表达情况。**结果** 共得到1254个DEGs, 其中表达上调基因422个, 表达下调基因832个。GO富集分析结果显示, DEGs主要与蛋白质异二聚体活性、阳离子跨膜转运蛋白活性、染色质结合等生物学功能相关; KEGG通路分析结果显示, DEGs主要富集在细胞周期、光传导、DNA复制等通路。共得到79个位于PPI网络中的核心基因。WGCNA分析将DEGs划分为11个共表达基因模块, 根据各个模块与临床性状之间的Pearson相关系数筛选出blue、pink、turquoise、red和brown共5个重要模块, 其中blue模块与诊断年龄的相关系数最高( $r=0.65$ ), 综合分析后得到染色体结构维持蛋白4(*smc4*)、微小染色体维持蛋白6(*mcm6*)、着丝粒蛋白K(*cenpk*)、驱动蛋白家族成员15(*kif15*)、胞质分裂蛋白调控因子-1(*prc1*)等5个关键基因。qRT-PCR验证结果表明, *prc1*、*cenpk*基因在Y79、WERI-RB-1细胞中的相对表达量明显高于ARPE-19细胞( $P<0.05$ ); 免疫组化检测结果显示, PRC1蛋白在RB组织样本中的表达量明显高于正常视网膜组织( $P<0.05$ )。**结论** 与RB发生相关的关键基因包括*smc4*、*mcm6*、*cenpk*、*kif15*、*prc1*等; PRC1蛋白在RB中表达量增高, 可能对RB的发生发展具有重要影响。

**[关键词]** 生物信息学; 视网膜母细胞瘤; 加权基因共表达网络分析; 差异表达基因

## Identification of the key genes in retinoblastoma based on bioinformatics analysis and experimental verification

Huang Zi-Shan, Fu Xin-Yu, Zhou Xi-Yuan\*

Department of Ophthalmology, the Second Affiliated Hospital of Chongqing Medical University/Chongqing Key Laboratory of Ophthalmology, Chongqing 400010, China

\*Corresponding author, E-mail: zhouxiyuan2002@aliyun.com

This work was supported by the National Natural Science Foundation of China (82070976), and the Key Sci-Tech Innovation Project for Social Undertakings and People's Livelihood in Chongqing (cstc2017shms-zdyfX0021)

[基金项目] 国家自然科学基金(82070976); 重庆市社会事业与民生保障科技创新专项重点研发项目(cstc2017shms-zdyfX0021)

[作者简介] 黄子珊, 硕士研究生, 主要从事眼底疾病方面的研究

[通信作者] 周希媛, E-mail: zhouxiyuan2002@aliyun.com

**[Abstract] Objective** To search for genes related to the occurrence and development of retinoblastoma (RB). **Methods** In this study, firstly, we obtained 3 gene expression datasets from the Gene Expression Omnibus (GEO) database. After they were merged, the sva package in R software was applied to remove the batch effects of these three datasets. The differentially expressed genes (DEGs) were identified by limma package. ClusterProfiler package was used to analyze GO enrichment and KEGG pathway of DEGs. STRING database and Cytoscape software were used to construct the protein-protein interaction network (PPI). CytoHubba was applied to find the hub gene of PPI network. Weighted gene co-expression network analysis (WGCNA) was utilized to identify key modules associated with clinical information. The key genes of the key modules were further searched. Then, Y79, WERI-RB-1 and ARPE-19 cells were cultured *in vitro*, and the mRNA expression levels of five key genes were detected by quantitative real-time polymerase chain reaction (qRT-PCR). A total of 20 RB tissue samples were collected. Immunohistochemistry was used to detect the expression of key proteins. **Results** A total of 1254 DEGs were identified, among which 422 were up-regulated and 832 were down-regulated. In GO analysis, DEGs were mainly related to protein heterodimerization activity, cation transmembrane transporter activity, and chromatin binding. In KEGG analysis, DEGs were mainly enriched in cell cycle, phototransduction, and DNA replication. A total of 79 hub genes in the PPI network were obtained. In the co-expression network, DEGs were divided into 11 co-expressed gene modules. According to Pearson correlation coefficient between each module and clinical traits, 5 important modules including blue, pink, turquoise, red and brown were identified, among which blue module had the highest correlation coefficient with age at diagnosis ( $r=0.65$ ). After comprehensive analysis, we obtained 5 hub genes including structural maintenance of chromosome 4 (*smc4*), minichromosome maintenance complex component 6 (*mcm6*), centromere protein K (*cenpk*), kinesin family member 15 (*kif15*), protein regulator of cytokinesis 1 (*prc1*). The qRT-PCR results showed that the mRNA relative expression levels of *prc1* and *cenpk* in Y79/WERI-RB-1 cells were higher than those in ARPE-19 ( $P<0.05$ ). Immunohistochemistry results indicated that PRC1 protein was higher expressed in RB tissue samples ( $P<0.05$ ). **Conclusion** *smc4*, *mcm6*, *cenpk*, *kif15*, and *prc1* were identified as hub genes by bioinformatics analysis, and the increased expression of PRC1 protein in RB may play an important role in the occurrence and development of RB.

**[Key words]** bioinformatics analysis; retinoblastoma; weighted gene co-expression network analysis; differentially expressed gene

视网膜母细胞瘤(retinoblastoma, RB)是婴幼儿最常见的眼内恶性肿瘤<sup>[1]</sup>,年发病率为1/18 000~1/16 000,可分为遗传型(约45%)和非遗传型(约55%)<sup>[2]</sup>,如果不及时治疗可能导致失明或死亡。为实现RB的早诊断、早治疗及降低病死率,寻找RB的关键基因尤为重要。加权基因共表达网络(weighted gene co-expression network analysis, WGCNA)分析目前已广泛应用于与肿瘤相关的高通量数据挖掘中<sup>[3-4]</sup>,该分析方法可寻找协同表达的基因模块,探索基因模块与关注的表型(如临床数据)之间的关系,构建共表达网络并寻找网络中的核心基因,主要通过基因之间相关系数计算、基因模块的确定、共表达网络、模块与性状关联4个步骤来寻找影响疾病的关键基因。本研究运用多种生物信息学分析方法寻找与RB发生发展相关的关键基因,并通过实时荧光定量PCR(quantitative real-time polymerase chain reaction, qRT-PCR)及免疫组化等多种方法对筛选出来的关键基因进行验证,以进一步寻找能够用于早期诊断RB的生物标志物。

## 1 材料与方法

**1.1 数据下载及处理** 从GEO(<https://www.ncbi.nlm.nih.gov/gds>)基因表达数据库中下载基因表达谱数据集GSE110811、GSE24673、GSE97508,3个芯

片数据集共包含43例RB样本和7例正常对照组织样本。其中,GSE110811基于GPL16686平台,包含28例RB和3例正常视网膜组织样本(本研究排除了1例正常对照样本GSM3017153,原因为该数据集样本可能混有肿瘤组织<sup>[5]</sup>);GSE24673基于GPL6244平台,包含9例RB和2例健康成人视网膜组织样本;GSE97508基于GPL15207平台,包含6例RB和3例正常视网膜组织样本。

根据平台注释信息将阵列探针名转换为匹配的基因名。采用perl软件将这3个数据集合并,用R软件的sva包消除3个数据集之间的批次效应(样品在不同时间、不同分组或由不同的人处理造成的差异)<sup>[6]</sup>。将3个数据集的矩阵数据合并为1个数据集,共包含43例RB和7例正常对照组织样本。

**1.2 筛选差异表达基因(differentially expressed gene, DEGs)** 使用R语言的limma包分析合并数据集中RB与正常对照样本之间的DEGs矩阵,筛选条件设置为 $|\log FC|>1$ , $P<0.05$ <sup>[7]</sup>。

**1.3 GO(gene ontology)富集分析和KEGG(Kyoto Encyclopedia of Genes and Genomes)通路分析** 对DEGs进一步分析,利用R语言clusterProfiler<sup>[8]</sup>包进行GO富集分析和KEGG通路分析。校正后的P值(FDR) $<0.05$ 为差异有统计学意义。

**1.4 蛋白相互作用(protein-protein interaction, PPI)**

网络 将DEGs输入STRING数据库,以交互得分=0.9(最高置信度)作为阈值构建PPI网络,然后导入Cytoscape(3.7.2版本)软件中进行可视化分析和整理,运用cytoHubba插件以最大集团拓扑分析法(maximal clique centrality, MCC)<sup>[9]</sup>选取关键节点的基因<sup>[10]</sup>。

**1.5 WGCNA分析** 提取GSE110811肿瘤组28个样本对应的DEGs矩阵进行WGCNA分析,获取该肿瘤样本对应的临床信息。

**1.5.1 共表达网络的构建与模块划分** 利用R软件的WGCNA包<sup>[11]</sup>,计算各基因间的Pearson相关系数,选择适当的软阈值 $\beta$ 构建无尺度网络;建立邻接矩阵,转换为TOM重叠矩阵;表达模式相似的基因分为一类,并构成模块;基于TOM矩阵,使用平均连锁层次聚类方法来聚类基因,根据混合动态切割树的标准设置最小模块,构建加权共表达网络以筛选hub基因<sup>[12]</sup>,以0.5作为加权共表达相关系数的阈值,导入Cytoscape(3.7.2版本)中进行可视化分析,找到位于网络中心的核心基因。

**1.5.2 关键模块筛选** 将样本的临床信息与模块基因进行关联分析,寻找显著相关的模块。本研究患者临床信息包括:有无远处转移、有无继发性肿瘤、RB1基因突变、单双侧眼患病、诊断年龄和随访时间等。采用Pearson相关系数评估模块与临床信息之间的相关性,相关系数越大表示该模块与临床数据相关程度越高,该模块即为关键模块,并计算各个基因与临床信息的相关系数。

**1.5.3 关键模块基因GO和KEGG分析** 使用DAVID数据库对关键模块的基因行GO富集分析和KEGG通路分析( $P < 0.05$ 为显著性基因富集的筛选条件)。

**1.6 实时荧光定量PCR(qRT-PCR)检测关键基因的表达量**

**1.6.1 细胞培养及分组** 人RB细胞株(Y79、WERI-RB-1)分别购自上海佰晔生物科技中心和上海富衡生物科技有限公司;人视网膜色素上皮细胞株(ARPE-19)购自上海富衡公司。Y79及WERI-RB-1细胞培养基均利用RPMI 1640培养基(美国Hyclone公司)加20%胎牛血清(上海依科赛生物)及1.0%青链霉素溶液(上海碧云天公司)配制,ARPE-19细胞培养基为含10%胎牛血清(上海依科赛生物)及1.0%青链霉素溶液的DMEM-F12(美国Hyclone公司)培养基;所有细胞均在37℃、5%CO<sub>2</sub>培养箱中培养。Y79、WERI-RB-1细胞作为实验组,ARPE-19细胞作为对照组。

**1.6.2 qRT-PCR实验步骤** 收集细胞,PBS洗2或3次,加入Trizol裂解液(cat: 15596026,美国Invitrogen公司),依次使用氯仿、异丙醇和75%乙醇提取总RNA;利用超微量分光光度计对RNA样

品进行浓度和纯度测定,利用反转录试剂盒(日本TaKaRa公司)并按照操作说明书配制反应液,分两步进行实验将RNA反转录为cDNA;使用SYBR Green qPCR Master Mix试剂盒(MCE公司)配制反应体系,依次加入无酶水、SYBR Green染料、上下游引物、cDNA模板后混匀,用ABI 7500荧光定量PCR仪进行扩增反应。设置热循环条件为:95℃ 5 min; 95℃ 10 s、60℃ 30 s、95℃ 15 s、60℃ 1 min、95℃ 30 s、60℃ 15 s,共40个循环<sup>[13]</sup>。以 $\beta$ -actin作为内参照,5个关键基因引物序列见表1(由上海生工有限公司合成)。实验重复3次。采用 $2^{-\Delta\Delta Ct}$ 法计算目的基因的相对表达量。

表1 qRT-PCR引物序列

Tab.1 Primer sequences for qRT-PCR

基因	引物序列(5'-3')
<i>smc4</i>	正义: CGCCTCCAGCAATGACCAAT
	反义: CCCAGCATAGGATTTGAAGTT
<i>mcm6</i>	正义: GAGGAACTGATTCGTCTGAGA
	反义: CAAGGCCCGACACAGGTAAG
<i>cenpk</i>	正义: AACACTCACCGATTCAAATGCT
	反义: CAGTCAAGGGAATTGTTTCAGGT
<i>kif15</i>	正义: AAAACTGAGTTACGCAGCGTG
	反义: AGTTGCGAATACAGATTCCTGAG
<i>prc1</i>	正义: ATCACCTTCGGGAAATATGGGA
	反义: TCTTTCTGACAGACGGATATGCT
$\beta$ -actin	正义: AGCGAGCATCCCCAAAGTT
	反义: GGGCACGAAGGCTCATCATT

**1.7 免疫组化检测关键蛋白表达量**

**1.7.1 临床标本收集及分组** 纳入2016年1月—2019年1月重庆医科大学附属第二医院收治的RB患者石蜡组织标本20例,所有患者术前均未进行放疗,获取10例因车祸伤等原因进行眼球摘除患者的正常视网膜作为对照组。本研究经重庆医科大学附属第二医院伦理委员会审批[(2019)270]。

**1.7.2 免疫组化实验步骤** 石蜡切片常规脱蜡后,将组织切片置于乙二胺四乙酸(EDTA)抗原修复缓冲液(pH9.0)中进行抗原修复,放入3% H<sub>2</sub>O<sub>2</sub>溶液避光孵育25 min;滴加3%牛血清白蛋白(BSA)室温封闭30 min;分别加入着丝粒蛋白K(CENPK)抗体(北京博奥森公司)、兔抗人胞质分裂蛋白调控因子-1(PRC1)抗体(Abcam公司)4℃孵育过夜<sup>[14]</sup>,加入辣根过氧化物酶(HRP)标记的山羊抗兔二抗;用3,3'-二氨基联苯胺(DAB)显色、苏木精复染,脱水后中性树脂封片,组织切片在显微镜下进行检查。免疫组化的结果分级标准(0~12分):0~4分为低表达组(阴性),5~12分为高表达组(阳性)<sup>[15]</sup>。上述试

剂除一抗外均购买自武汉赛维尔生物公司。

**1.8 统计学处理** 采用GraphPad Prism 8.3.0软件进行qRT-PCR结果作图及统计分析,多组间关键基因表达量的比较采用单因素方差分析,进一步两两比较采用Dunnnett-*t*检验;采用SPSS 25.0软件对免疫组化结果进行分析, RB组织与对照组中关键蛋白的表达差异采用单侧Fisher确切概率法进行比较。 $P < 0.05$ 为差异有统计学意义。

## 2 结果

**2.1 DEGs的筛选** 共得到1254个DEGs,其中表达上调基因422个,表达下调基因832个,具体分布情况见图1。

**2.2 DEGs的功能注释和通路富集** GO富集分析结果显示,DEGs主要与蛋白质异二聚体活性、阳离子跨膜转运蛋白活性、染色质结合、无机阳离子跨膜转运活性和细胞黏附分子结合等生物学功能密切相关(图2A)。KEGG通路分析结果显示,DEGs主要富集在细胞周期、光传导、DNA复制和卵母细胞减数分裂等通路上(图2B)。

**2.3 PPI网络构建及核心基因筛选** PPI网络由3567条边、611个节点构成,以MCC $> 2.09 \times 10^{13}$ 为标准得到79个基因<sup>[9]</sup>,即为网络核心基因(图3)。

### 2.4 WGCNA分析结果

**2.4.1 共表达网络构建与枢纽基因筛选** 利用pickSoftThreshold函数筛选合适的软阈值 $\beta$ ,使无尺度拓扑拟合指数 $R^2 > 0.8$ 以构建共表达网络(图4A), $\beta = 6$ 时共表达网络中基因之间的连接性满足无标度

网络分布(图4B),利用MCC算法以得分 $> 4$ 作为标准<sup>[9]</sup>,得到位于网络中心的前13个基因,即为枢纽基因(图4C)。

**2.4.2 基因模块的划分与关键模块筛选** 利用动态混合剪切法获得11个共表达基因模块,其中灰色模块是无法聚集到其他模块的基因,基于模块基因构建模块聚类树(图5A),模块的拓扑重叠热图反映模块与模块间基因的相关程度(图5B),基因的聚类树状图展示模块划分过程(图5C)。分别计算各个模块与临床性状之间的Pearson相关系数及P值(图6A),以Pearson相关系数的绝对值 $> 0.55$ 且 $P < 0.05$ 作为筛选条件,筛选出5个重要模块,分别是blue、pink、turquoise、red和brown模块。Blue、pink和turquoise模块与诊断年龄呈明显正相关( $r = 0.65, P = 2 \times 10^{-4}$ ;  $r = 0.58, P = 0.001$ ;  $r = 0.59, P = 0.001$ ),其中pink模块还与单双侧眼患病和随访时间呈明显负相关( $r = -0.57, P = 0.002$ ;  $r = -0.63, P = 4 \times 10^{-4}$ );而red和brown模块与继发性肿瘤呈明显正相关( $r = 0.57, P = 0.001$ ;  $r = 0.55, P = 0.002$ )。其中blue模块与诊断年龄的相关系数最高,将blue模块包含的基因导入STRING数据库构建PPI网络(图6B)。各个重要模块包含的基因见表2。本研究还将基因与临床性状关联,分别计算出每个基因与临床性状的相关性,选取与每项临床性状相关系数绝对值最高的基因,得到hmgb3、pcdhb3、ids、polh、birc5、guca1b共6个基因(表3),其中pcdhb3与诊断年龄的相关系数最高( $r = 0.8411$ )。

**2.4.3 GO富集分析和KEGG通路分析** Pink和red

表2 视网膜母细胞瘤研究中各个重要模块包含的基因

Tab.2 Genes contained in each important module of research on RB

模块	基因
Blue	<i>tpx2, kif11, aspm, ndc80, ccnb2, arhgap11a, dlgap5, bub1, plk4, depdc1, cenpe, kif18a, smc4, kif20b, mad211</i>
Brown	<i>hells, mom6, nusap1, dtl, gjc1, ncapp2, prc1, ezh2, wdr76, kntc1, fanci, hmgb2, nasp</i>
Turquoise	<i>top2a, mki67, chek1, ncapp, cenpf, cenpk, ckap2l, kif15, kif23, rtkn2, lmnbl, kiaa1524, dna2, tmpo, atad5, smc2</i>
Pink	<i>mis18bp1, stil</i>
Red	<i>atad2, smc3</i>

表3 视网膜母细胞瘤与临床性状相关程度较高的基因列表

Tab.3 The genes with the higher degree of correlation for each clinical trait in RB

临床性状	基因	相关系数
远处转移	<i>guca1b</i>	0.6248
继发性肿瘤	<i>polh</i>	0.6935
RB1基因突变	<i>birc5</i>	0.6145
单双侧眼患病	<i>hmgb3</i>	-0.6813
诊断年龄	<i>pcdhb3</i>	0.8411
随访时间	<i>ids</i>	-0.6920

模块包含基因数目较少,无法进一步进行富集分析,因此将blue、turquoise和brown模块中的基因导入DAVID数据库中,以校正后的P值(FDR) $< 0.05$ 作为筛选条件。Blue模块GO分析结果显示,对于生物过程主要富集在细胞分裂、有丝核分裂等过程,对于细胞组分主要富集在细胞核、细胞质、微管等结构,对于分子功能主要富集在ATP结合、微管结合等功能方面;KEGG分析结果显示主要富集在细胞周期、卵母细胞减数分裂等通路。Turquoise模块基因主要与细胞核、核质、ATP结合等相关,

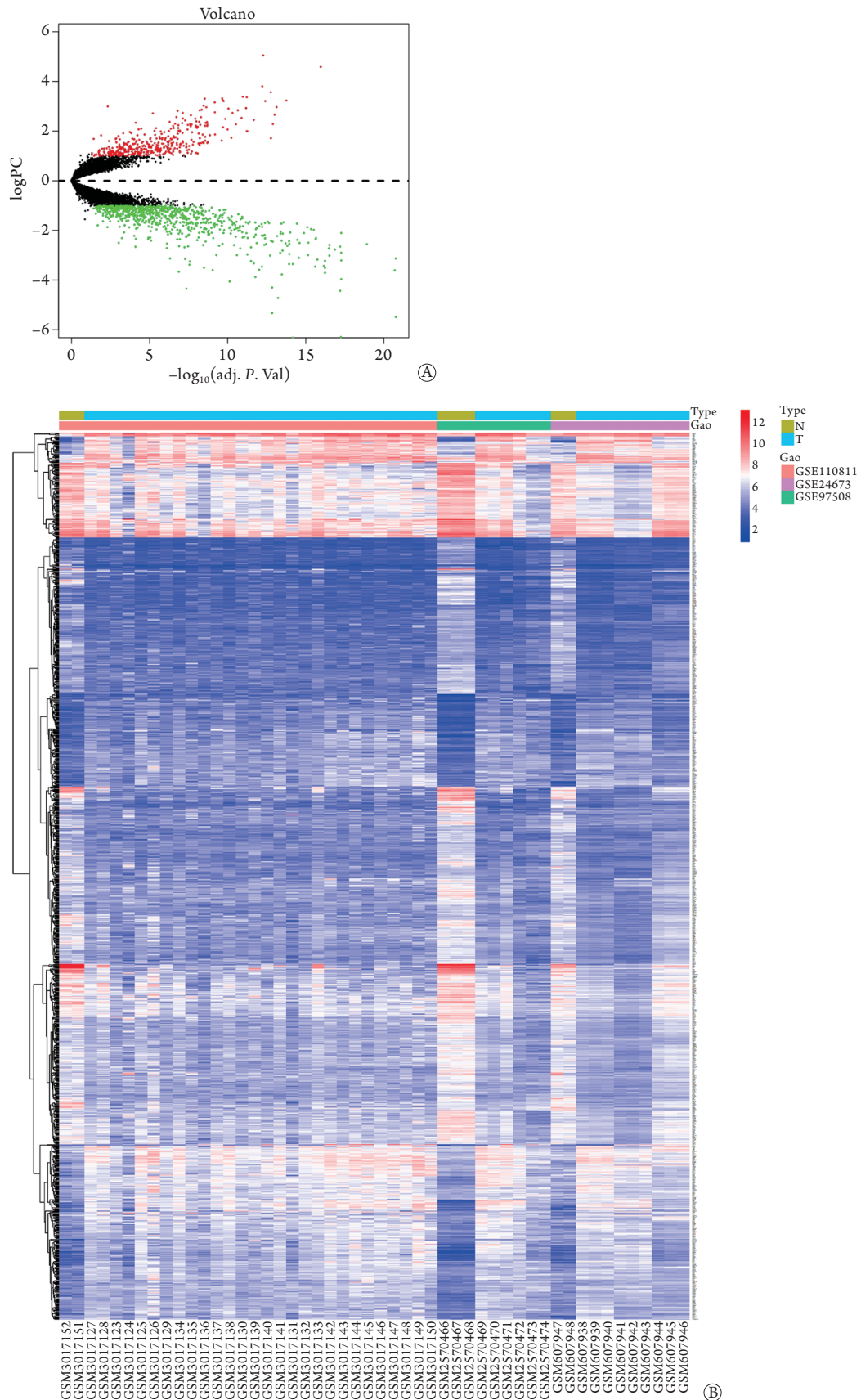


图1 视网膜母细胞瘤组织与正常组织差异表达基因的火山图及热图

Fig.1 Volcano and heat maps of differentially expressed genes in RB tissues and normal tissues

A. 火山图(红色点表示表达上调基因; 绿色点表示表达下调基因); B. 差异表达聚类热图(蓝色表示表达下调基因, 红色表示表达上调基因); 顶部Type横条: 黄绿色表示正常组织样本, 蓝色表示RB组织样本; Geo横条: 分别标注了3个数据集包含的组织样本

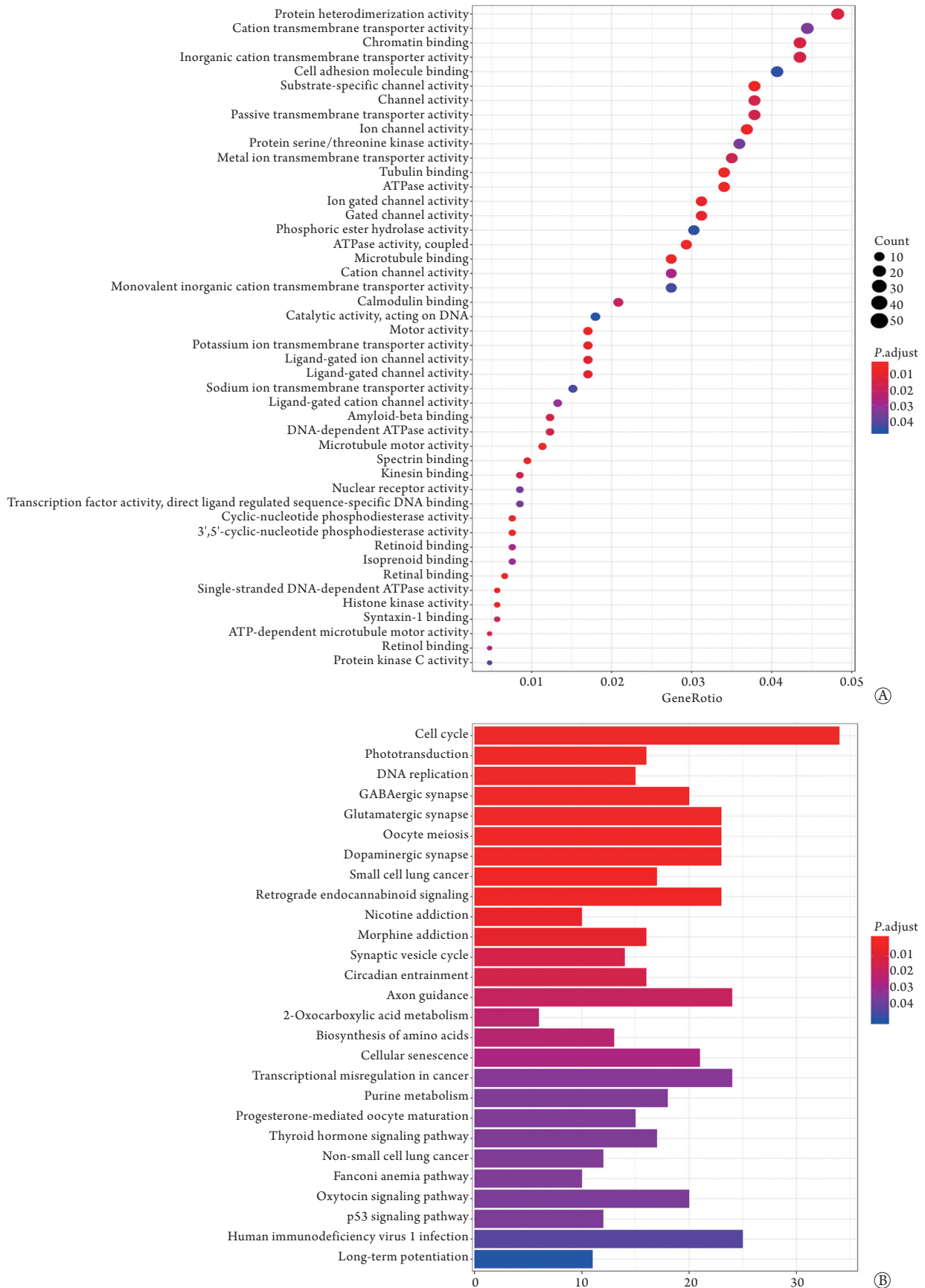


图2 视网膜母细胞瘤差异表达基因的GO富集分析(A)和KEGG通路分析(B)

Fig.2 GO enrichment (A) and KEGG pathway analysis (B) of differentially expressed genes in RB

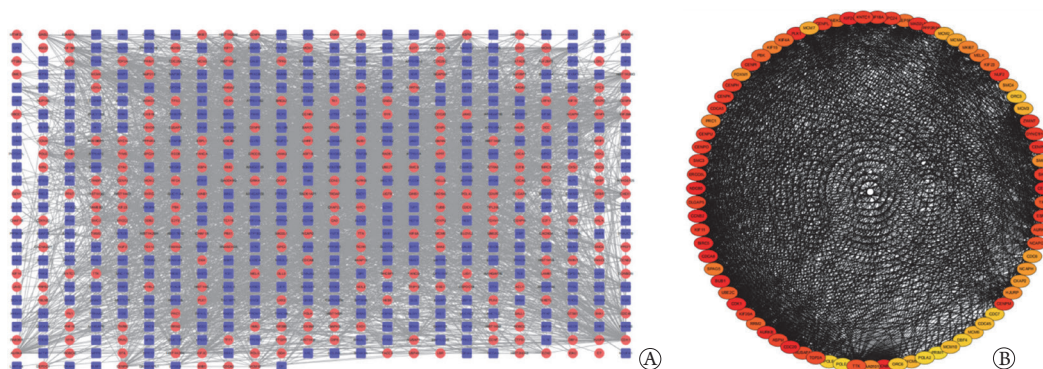


图3 视网膜母细胞瘤差异表达基因PPI网络构建及核心基因筛选

Fig.3 PPI networks of DEGs and screening core genes in RB

PPI. 蛋白相互作用; DEG. 差异表达基因; A. PPI网络(利用Cytoscape软件对DEGs构建网络图, 红色圆形表示上调基因, 紫色正方形表示下调基因); B. 网络核心基因(cytoHubba中MCC算法得分从高到低排序得到79个基因, 颜色越深表示得分越高)

brown模块基因主要富集在细胞核、核质等细胞成分上。

**2.5 关键基因筛选** 将DEGs PPI的核心基因、WGCNA网络的枢纽基因及重要模块包含基因取交集(图6C), 得到染色体结构维持蛋白4(structural maintenance of chromosome 4, *smc4*)、微小染色体维持蛋白6(minichromosome maintenance complex component 6, *mcm6*)、着丝粒蛋白K(centromere protein K, *cenpk*)、驱动蛋白家族成员15(kinesin family member 15, *kif15*)、胞质分裂蛋白调控因子-1(protein regulator of cytokinesis 1, *prc1*)等5个关键基因。上述5个关键基因在RB中均为高表达基因, 其logFC均>1, 关键基因所属模块及差异表达信息见表4。

表4 视网膜母细胞瘤关键基因所属模块及差异表达信息  
Tab.4 The modules and differential expression information of key genes in RB

关键基因	模块	logFC	校正后P值
<i>smc4</i>	Blue	1.4902	<0.01
<i>mcm6</i>	Brown	2.6596	<0.01
<i>cenpk</i>	Turquoise	2.9585	<0.01
<i>kif15</i>	Turquoise	2.3706	<0.01
<i>prc1</i>	Brown	2.2593	<0.01

*smc4*. 染色体结构维持蛋白4; *mcm6*. 微小染色体维持蛋白6; *cenpk*. 着丝粒蛋白K; *kif15*. 驱动蛋白家族成员15; *prc1*. 胞质分裂蛋白调控因子-1

**2.6 关键基因在细胞株中的表达情况** 采用qRT-PCR检测5个关键基因的相对表达量, 结果显示, *prc1*和*cenpk*的相对表达量在Y79、WERI-RB-1细胞中明显高于对照组( $P<0.05$ , 图7)。

**2.7 RB及正常视网膜组织中PRC1、CENPK蛋白的表达情况** 免疫组化结果表明, 20例RB中有10例

(50%)PRC1蛋白表达阳性, 而10例正常视网膜中仅有1例(10%)呈阳性, 故RB组织中PRC1蛋白阳性率高于正常视网膜, 且差异有统计学意义( $P<0.05$ )。CENPK蛋白在RB和正常视网膜中的表达差异无统计学意义( $P>0.05$ , 图8)。

### 3 讨论

RB发生率居小儿恶性肿瘤的第2位<sup>[1]</sup>, 严重威胁患儿生命。RB发病多在5岁以下, 约2/3的患儿在3岁前患病, 近年来其发病率呈上升趋势, 且如果治疗不及时, 可能导致失明或死亡。因此, 为实现RB的早诊断、早治疗, 寻找其关键基因及生物标志物显得越来越重要。

目前, 高通量数据资源越来越丰富, 许多学者运用多种生物信息学方法进行大样本数据挖掘<sup>[16]</sup>, 寻找导致某种疾病发生的关键基因, 以进一步发现与疾病诊断及预后相关的生物标志物, 此方式较传统研究方法更省时省力、节约经费且科学有效。然而, 以往的生物信息学分析多使用单个数据集<sup>[17]</sup>和单一方法分析DEGs<sup>[18]</sup>, 而RB发病率较低且组织样本获取难度大, 故组织样本较为稀缺, 单个数据集样本量较少, 可能导致分析结果不准确。因此, 本研究将3个RB数据集中的原始数据进行合并以增加样本量, 运用差异表达分析联合加权基因共表达网络分析, 将DEGs划分成不同模块并与临床性状相关联, 筛选出了与RB发生发展密切相关的5个可能基因。其中, SMC4对哺乳动物的DNA修复至关重要, 可能参与了基因表达沉默状态的维持和DNA修复等过程, 与肺腺癌、神经胶质瘤及结直肠肿瘤发生相关, 且与乳腺癌预后相关<sup>[19-22]</sup>。MCM6是高度保守的微型染色体维持蛋白之一, 是启动真核基因组复制所必需的, Liu等<sup>[23]</sup>发现其可促进肝细胞癌转移, 且可作为肝细胞癌早期复发的一种新的血

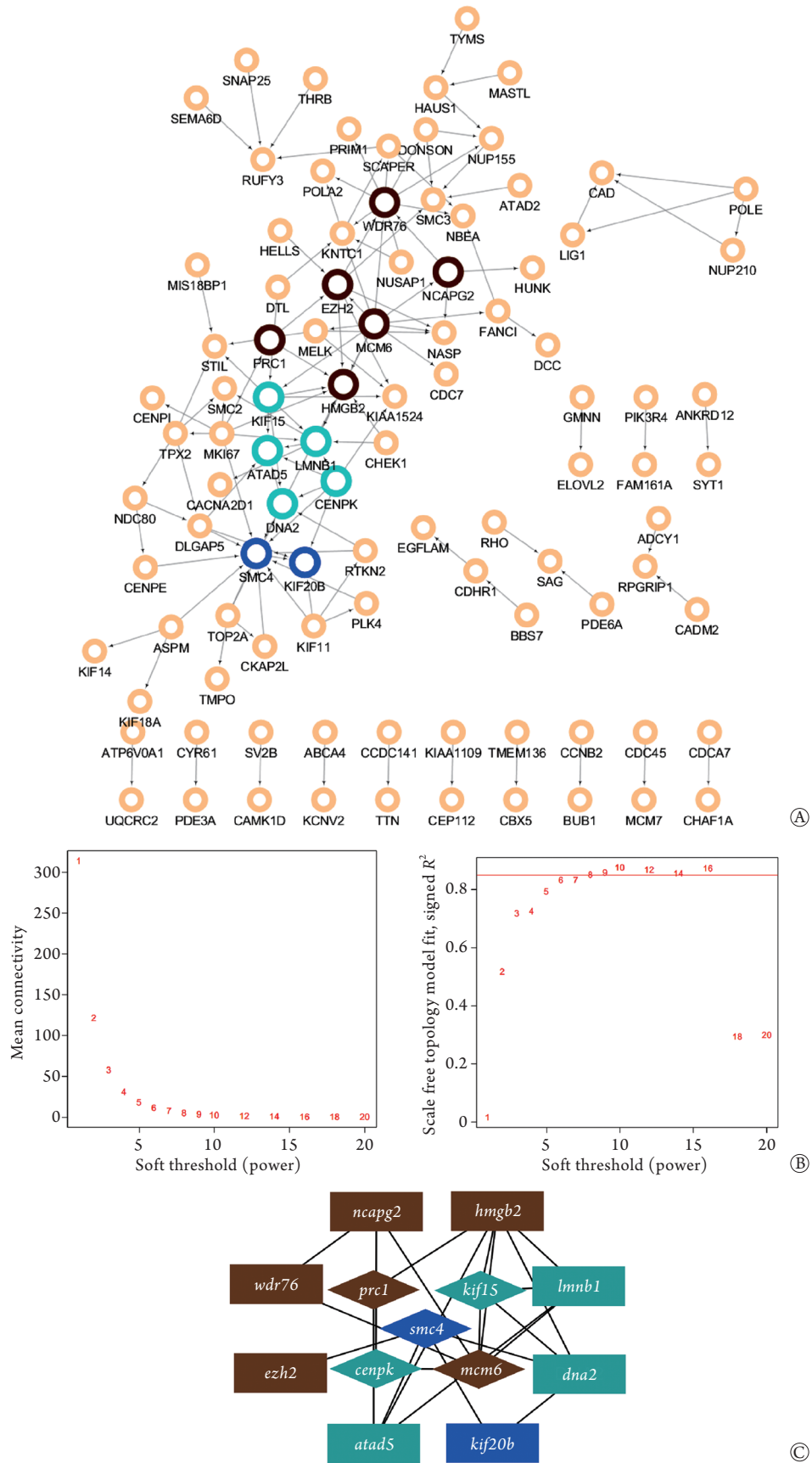


图4 视网膜母细胞瘤WGCNA网络构建及其枢纽基因的筛选

Fig.4 Screening of WGCNA network and key genes in RB

A. WGCNA网络(其中位于前13个的枢纽基因已突出显示, 不同颜色代表其所属模块); B. 软阈值 $\beta$ 筛选过程; C. WGCNA网络枢纽基因[蓝色、蓝绿色、棕色分别代表其所属模块(blue、turquoise、brown模块), 用菱形表示5个关键基因]

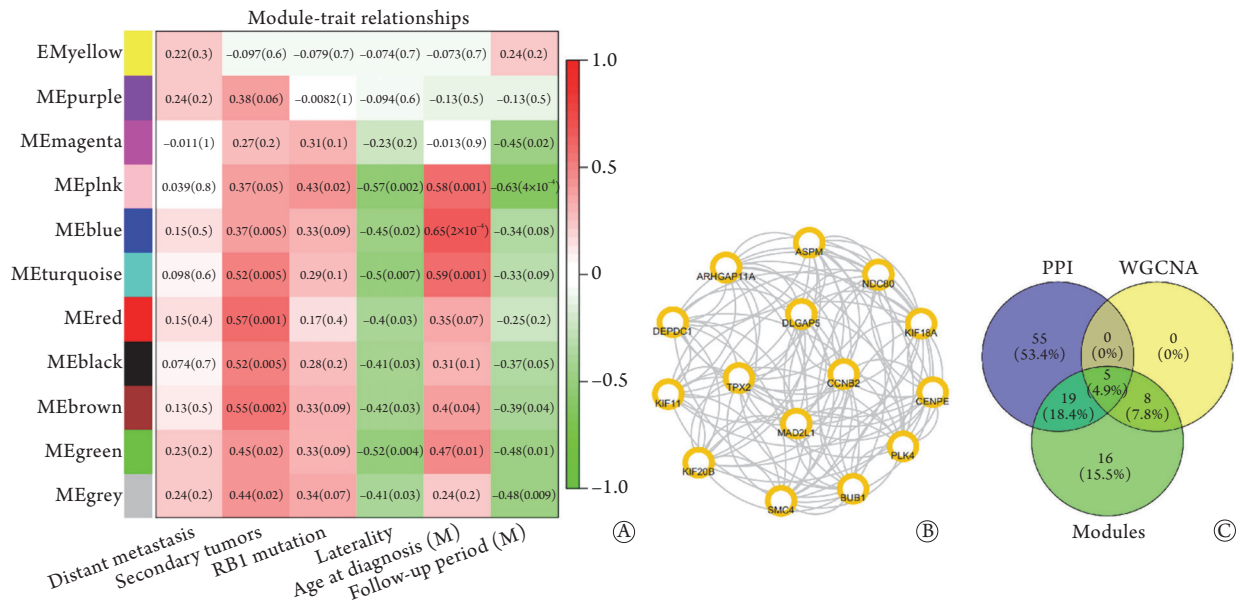


图5 视网膜母细胞瘤WGCNA分析模块划分过程

Fig.5 WGCNA analysis module partition process in RB

A. 模块聚类图; B. 模块的拓扑重叠热图(对角线颜色越深表示该模块拓扑重叠程度越高); C. 基因的聚类树状图(共获得11个共表达基因模块, 在底部用不同颜色表示)

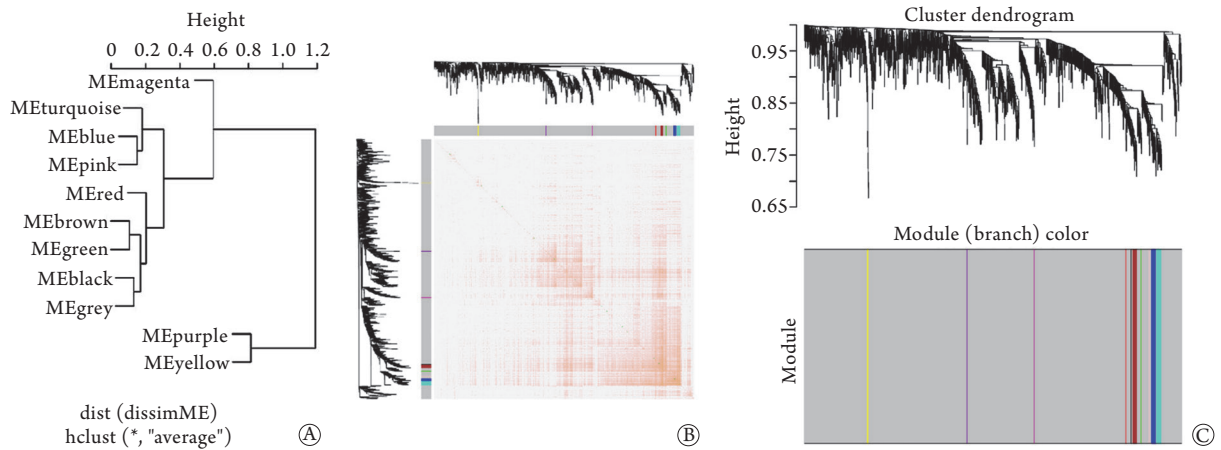


图6 模块与临床性状相关性热图及blue模块PPI网络

Fig.6 Heat map of correlation between module and clinical traits and PPI network of blue module

PPI. 蛋白相互作用; A. 模块与临床性状相关性热图(横轴表示临床性状, 左侧纵轴不同颜色表示不同模块, 每个格子中标明了相关系数, 括号内为对应P值, 颜色越深表示相关程度越大); B. blue模块基因的PPI网络; C. 5个关键基因筛选过程韦恩图

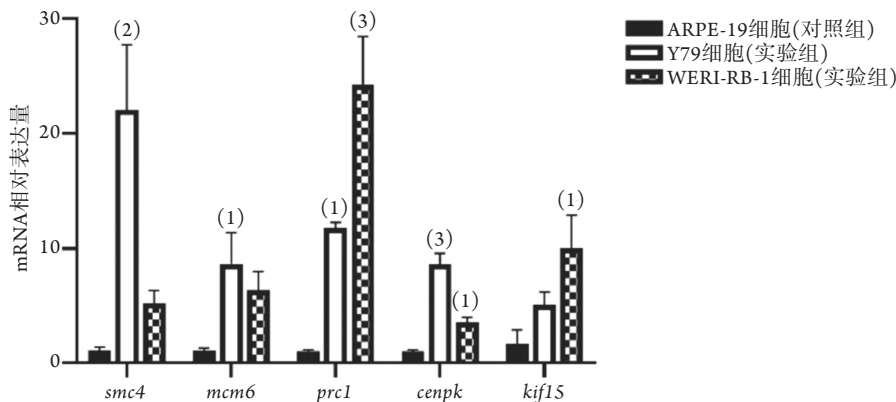


图7 视网膜母细胞瘤关键基因的相对表达量

Fig.7 Relative expression levels of key genes in RB

与ARPE-19细胞比较, (1) $P < 0.05$ , (2) $P < 0.01$ , (3) $P < 0.001$

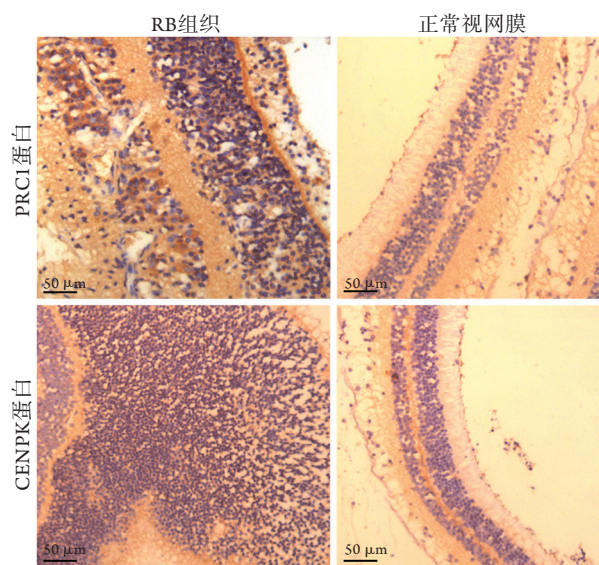


图8 PRC1及CENPK蛋白在视网膜母细胞瘤组织及正常视网膜中的表达(免疫组化染色,  $\times 200$ )

Fig.8 Expression of PRC1 and CENPK protein in RB tissue and normal retina (Immunohistochemical staining,  $\times 200$ )

清生物标志物<sup>[24]</sup>。CENPK与着丝粒功能和有丝分裂进程密切相关,有研究发现,它可能参与了肝细胞癌的恶性化进展<sup>[25]</sup>。Lee等<sup>[26]</sup>发现,CENPK在卵巢癌细胞中表达上调且可作为卵巢癌的一种新的肿瘤标志物。KIF15可促进黑色素瘤生成<sup>[27]</sup>,也是促进胰腺癌增殖的重要调控因子<sup>[28]</sup>。PRC1基因位于15q26.1,与细胞增殖、胞质分裂、纺锤体及微管的形成密切相关。

本研究的生物信息学分析结果显示,这5个基因均为RB组织中表达上调的基因。WGCNA分析结果显示它们均为重要模块中的基因,其中*smc4*属于blue模块,*mcm6*、*prc1*属于brown模块,*cenpk*、*kif15*则是turquoise模块的组成部分。由于blue模块与诊断年龄呈明显正相关,brown及turquoise模块与继发性肿瘤呈明显正相关,因此,*smc4*可能与RB的发病进程相关,而*mcm6*、*prc1*、*cenpk*、*kif15*可能与RB继发性肿瘤的发生存在一定相关性。但由于目前这些基因在RB中的研究较少,其在RB中的功能仍需进一步探索。另外,通过对上述关键模块基因的GO及KEGG分析发现,此类基因主要参与细胞核、微管等结构形成,与细胞周期及有丝分裂过程等通路密切相关,因此,推测这些关键模块中的差异表达基因可能是通过调控细胞生物学过程及信号通路来影响RB的发病进程的。此外,本研究还将基因与临床性状关联,分别计算出每个基因与临床性状的相关性,选取与每项临床性状相关系数绝对值最高的基因,得到了*pcdhb3*等共6个基因,其中*pcdhb3*与临床性状(诊断年龄)的相关系数最高。

有研究发现,PCDHB3是一种新型的结直肠癌肿瘤抑制因子,且可能是晚期结直肠癌的预后标志物<sup>[29]</sup>,因此,此类基因在RB中的功能也值得进一步研究。

为验证上述筛选出的关键基因是否在RB细胞株中高表达,本研究体外培养了Y79、WERI-RB-1及ARPE-19细胞进行qRT-PCR实验,结果表明*prc1*及*cenpk*这两个基因在RB细胞中高表达( $P < 0.05$ )。此外,通过前面的生物信息学分析可以发现,在这5个基因中*cenpk*、*prc1*在RB中的表达明显上调,而且它们在肿瘤的发生中发挥重要作用。结合qRT-PCR实验结果,我们挑选*cenpk*、*prc1*这两个基因进行进一步实验,即在RB组织样本中利用免疫组化实验在蛋白质水平进行验证,结果表明PRC1蛋白在RB组织中的表达量明显增高( $P < 0.05$ )。PRC1在恶性肿瘤中的作用研究尚少,Wang等<sup>[30]</sup>发现,PRC1高表达在肝癌细胞中增加了癌细胞的化疗耐药性。Liao等<sup>[31]</sup>对2株RB细胞(HXO-RB44和WERI-Rb-1)进行siRNA转染使*prc1*表达量降低后,发现RB细胞的增殖和血管生成受到了抑制,该过程是通过抑制Wnt/ $\beta$ -catenin通路实现的,表明PRC1蛋白对RB的发生具有重要影响。

综上所述,本研究运用差异表达分析联合加权基因共表达网络分析提取出了与RB相关的关键基因,其中RB细胞及组织中的PRC1表达量明显高于对照组,且Liao等<sup>[31]</sup>的体外实验表明沉默*prc1*基因可抑制RB细胞的增殖和血管生成,这些结果均提示PRC1可能对RB的发生发展具有一定的促进作用。此外,本研究对探究RB的发病机制及潜在的治疗靶点提供了新视角,但还需进行进一步体内实验来验证这些关键基因的生物功能。

#### 【参考文献】

- [1] Dimaras H, Kimani K, Dimba EAO, et al. Retinoblastoma[J]. Lancet, 2012, 379(9824): 1436-1446.
- [2] 陈大年, 范依萌. 关注RB1基因突变检测在视网膜母细胞瘤诊疗及遗传咨询中的作用[J]. 中华实验眼科杂志, 2018, 36(10): 742-747. [Chen DN, Fan YM. Emphasizing on the importance of RB1 genetic testing in diagnosis, treatment and genetic counseling of retinoblastoma[J]. Chin J Exp Ophthalmol, 2018, 36(10): 742-747.]
- [3] Tang XZ, Huang XL, Wang DP, et al. Identifying gene modules of thyroid cancer associated with pathological stage by weighted gene co-expression network analysis[J]. Gene, 2019, 704: 142-148.
- [4] Zhang L, Kang WQ, Lu XL, et al. Weighted gene co-expression network analysis and connectivity map identifies lovastatin as a treatment option of gastric cancer by inhibiting HDAC2[J]. Gene, 2019, 681: 15-25.
- [5] Hudson LE, Mendoza P, Hudson WH, et al. Distinct gene expression profiles define anaplastic grade in retinoblastoma[J].

- Am J Pathol, 2018, 188(10): 2328-2338.
- [6] Leek JT, Johnson WE, Parker HS, *et al.* The sva package for removing batch effects and other unwanted variation in high-throughput experiments[J]. *Bioinformatics*, 2012, 28(6): 882-883.
- [7] Wang S, Yu ZH, Chai KQ. Identification of CFTR as a novel key gene in chromophobe renal cell carcinoma through bioinformatics analysis[J]. *Oncol Lett*, 2019, 18(2): 1767-1774.
- [8] Yu G, Wang L, Han Y, *et al.* clusterProfiler: an R package for comparing biological themes among gene clusters[J]. *OMICS*, 2012, 16(5): 284-287.
- [9] Zhang Y, Yang X, Zhu XL, *et al.* Bioinformatics analysis of potential core genes for glioblastoma[J]. *Biosci Rep*, 2020, 40(7): BSR20201625.
- [10] 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.
- [11] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis[J]. *BMC Bioinformatics*, 2008, 9: 559.
- [12] Chen S, Yang D, Lei C, *et al.* Identification of crucial genes in abdominal aortic aneurysm by WGCNA[J]. *Peer J*, 2019, 7: e7873.
- [13] Wu XZ, Cui HP, Lv HJ, *et al.* Knockdown of lncRNA PVT1 inhibits retinoblastoma progression by sponging miR-488-3p[J]. *Biomedicine Pharmacother*, 2019, 112: 108627.
- [14] 步华磊. PRC1在高级别浆液性卵巢癌中生物学功能与临床特征研究[D]. 济南: 山东大学, 2020. [Bu HL. Biological function and clinical characteristics of PRC1 in high-grade serous ovarian cancer[D]. Jinan: Shandong University, 2020.]
- [15] 齐贵胜. PCNA、VEGF和E-cad在乳腺浸润性导管癌组织中的表达及分析[D]. 济南: 山东师范大学, 2010. [Qi GS. Expression and analysis of PCNA, VEGF and E-cad in invasive ductal carcinoma of breast[D]. Jinan: Shandong Normal University, 2010.]
- [16] 张钧栋, 杨波, 陈浩然, 等. 新型冠状病毒感染致造血损伤及潜在治疗药物的生物信息学分析[J]. *解放军医学杂志*, 2020, 45(9): 947-956. [Zhang JD, Yang B, Chen HR, *et al.* Bioinformatics analysis of hematopoietic injury and potential therapeutic drugs associated with SARS-CoV-2 infection[J]. *Med J Chin PLA*, 2020, 45(9): 947-956.]
- [17] Liu Z, Liu S, Guo J, *et al.* Identification and analysis of key genes driving gastric cancer through bioinformatics[J]. *Genet Test Mol Biomarkers*, 2021, 25(1): 1-11.
- [18] Guo S, Wu J, Zhou W, *et al.* Identification and analysis of key genes associated with acute myocardial infarction by integrated bioinformatics methods[J]. *Medicine (Baltimore)*, 2021, 100(15): e25553.
- [19] Zhang C, Kuang M, Li M, *et al.* SMC4, which is essentially involved in lung development, is associated with lung adenocarcinoma progression[J]. *Sci Rep*, 2016, 6: 34508.
- [20] Jiang L, Zhou J, Zhong D, *et al.* Overexpression of SMC4 activates TGF $\beta$ /Smad signaling and promotes aggressive phenotype in glioma cells[J]. *Oncogenesis*, 2017, 6(3): e301.
- [21] Feng XD, Song Q, Li CW, *et al.* Structural maintenance of chromosomes 4 is a predictor of survival and a novel therapeutic target in colorectal cancer[J]. *Asian Pac J Cancer Prev*, 2014, 15(21): 9459-9465.
- [22] Ma RM, Yang F, Huang DP, *et al.* The prognostic value of the expression of SMC4 mRNA in breast cancer[J]. *Dis Markers*, 2019, 2019: 2183057.
- [23] Liu M, Hu Q, Tu M, *et al.* MCM6 promotes metastasis of hepatocellular carcinoma via MEK/ERK pathway and serves as a novel serum biomarker for early recurrence[J]. *J Exp Clin Cancer Res*, 2018, 37(1): 10.
- [24] Liu Z, Li J, Chen J, *et al.* MCM family in HCC: MCM6 indicates adverse tumor features and poor outcomes and promotes S/G2 cell cycle progression[J]. *BMC Cancer*, 2018, 18(1): 200.
- [25] Wang J, Li H, Xia C, *et al.* Downregulation of CENPK suppresses hepatocellular carcinoma malignant progression through regulating YAP1[J]. *Onco Targets Ther*, 2019, 12: 869-882.
- [26] Lee YC, Huang CC, Lin DY, *et al.* Overexpression of centromere protein K (CENPK) in ovarian cancer is correlated with poor patient survival and associated with predictive and prognostic relevance[J]. *Peer J*, 2015, 3: e1386.
- [27] Yu X, He X, Heindl LM, *et al.* KIF15 plays a role in promoting the tumorigenicity of melanoma[J]. *Exp Eye Res*, 2019, 185: 107598.
- [28] Wang J, Guo X, Xie C, *et al.* KIF15 promotes pancreatic cancer proliferation via the MEK-ERK signalling pathway[J]. *Br J Cancer*, 2017, 117(2): 245-255.
- [29] Ye W, Ling SP, Liu RY, *et al.* Exome sequencing reveals the genetic landscape and frequent inactivation of PCDHB3 in Chinese rectal cancers[J]. *J Pathol*, 2018, 245(2): 222-234.
- [30] Wang Y, Shi F, Xing GH, *et al.* Protein regulator of cytokinesis PRC1 confers chemoresistance and predicts an unfavorable postoperative survival of hepatocellular carcinoma patients[J]. *J Cancer*, 2017, 8(5): 801-808.
- [31] Liao YJ, Yin XL, Deng Y, *et al.* PRC1 gene silencing inhibits proliferation, invasion, and angiogenesis of retinoblastoma cells through the inhibition of the Wnt/ $\beta$ -catenin signaling pathway[J]. *J Cell Biochem*, 2019, 120(10): 16840-16852.

(责任编辑: 熊晓然)