

doi: 10.3978/j.issn.1000-4432.2021.02.02

View this article at: <http://dx.doi.org/10.3978/j.issn.1000-4432.2021.02.02>

视网膜母细胞瘤基因表达谱芯片的生物信息学分析

张艳, 吴瑜瑜

(福建医科大学附属第二医院眼科, 福建 泉州 362000)

[摘要] 目的: 应用生物信息学方法分析视网膜母细胞瘤(retinoblastoma, RB)基因表达谱芯片, 从转录水平探讨RB与正常视网膜组织的差异表达基因(differentially expressed genes, DEGs)的功能及相互作用。方法: 通过公共基因表达数据库(gene expression omnibus, GEO)数据库中选取GSE97508和GSE110811两个芯片数据, 共有34个RB组织和6个正常视网膜组织。利用GEO2R工具对和Draw Venn Diagrams软件, 筛选出正常视网膜组织与RB组织间的DEGs。通过在线STRING检索工具(search tool for the retrieval of interacting genes, STRING), 对DEGs进行基因本体注释(gene ontology, GO)、KEGG富集通路分析、蛋白互作(protein-protein interaction, PPI)K分析。结果: 在2个芯片中发现视网膜母细胞瘤共有20个DEGs, 包括3个上调DEGs和17个下调DEGs, GO本体注释结果显示表达上调DEGs所富集的功能主要在细胞分裂、染色体浓缩、核分裂和DNA构象改变。下调DEGs主要富集在光传导、视觉感知、光感受器细胞修复、感光细胞内外节和调节视紫红质介导的信号通路等。KEGG通路显示上调DEGs没有显著的信号通路, 下调DEGs参与光传导信号通路, 其中包括CNGA1, CNGB1, RHO, SAG四个基因, 通过PPI网络提示这4个基因相互联系, 并发现与其他节点连接最紧密核心基因RHO。结论: 利用生物信息学方法能有效对RB基因芯片数据进行挖掘, 为进一步研究RB发生发展机制及寻找潜在的药物治疗靶点提供参考。

[关键词] 生物信息学分析; 差异表达基因; 视网膜母细胞瘤

Bioinformatics analysis of differentially expressed genes in retinoblastoma

ZHANG Yan, WU Yuyu

(Department of Ophthalmology, Second Affiliated Hospital of Fujian Medical University, Quanzhou Fujian 362000, China)

Abstract **Objective:** To find differentially expressed genes between retinoblastoma and normal retinal tissues by bioinformatics analysis, and to investigate their molecular function and interactions in the transcriptional level. **Methods:** The gene expression profile datasets GSE97508 and GSE110811, including 34 retinoblastoma (RB) tissues and 6 normal retinal tissues, were downloaded from gene expression omnibus (GEO). Differentially expressed genes (DEGs) between normal retinal tissues and RB tissues were identified by GEO2R tool and the

收稿日期 (Date of reception): 2020-12-03

通信作者 (Corresponding author): 吴瑜瑜, Email: wuyeyedoctor@126.com

Draw Venn Diagrams software. The gene ontology (GO) analysis, KEGG pathway and protein-protein interaction (PPI) were analyzed by STRING. **Results:** In two microarrays of retinoblastoma we found total 20 DEGs were identified, including 3 up-regulated and 17 down-regulated genes. The GO ontology annotation results showed that the enrichment functions of up-regulated genes were mainly in cell division, chromosome enrichment, nuclear division, and DNA conformation change. Down-regulated genes were mainly concentrated in light conduction, visual perception, photoreceptor cell repair, photoreceptor cell inner and outer segment, and regulation of rhodopsin mediated signaling pathway. The KEGG pathway showed that there was no significant signal pathway in which up-regulated genes up-regulated DEGs, and down-regulated genes were involved in the phototransduction signaling pathway, including four genes of *CNGA1*, *CNGB1*, *RHO* and *SAG*. PPI network suggested that these four genes were interlinked, and RHO was found to be the most closely connected core gene with other nodes. **Conclusion:** Bioinformatics can be used effectively to analyze RB microarray data to provide theoretical reference for further exploration of tumorigenesis mechanism and help search for potential drug therapeutic targets.

Keywords bioinformatics analysis; differentially expressed genes; retinoblastoma

视网膜母细胞瘤(retinoblastoma, RB)是儿童最常见眼内恶性肿瘤,可损害患儿视力、眼球,甚至危及生命。2/3患者3岁以内发病,在世界范围内,新生儿发病率为1:16 000~1:18 000,每年约有8 000例新生病例确诊^[1]。虽然已经发现*RB1*基因的突变或缺失是RB发生的主要原因,即使RB具有明确*RB1*突变,癌变发展过程中也会发生其他基因组的变化,例如转录相关基因(*BCOR*)是继*RB1*之后最常见的突变基因^[2]。而且有一部分单侧发病的病例未发现*RB1*突变,因此近年来在探索*RB1*突变之外的其他发病机制。基因芯片技术能快速检测差异表达基因(differentially expressed genes, DEGs),经过十余年的应用,已被证明是一种可靠的技术^[3],大量基因芯片上的数据存储在公共数据库中,通过这些数据库,整合生物信息学方法对RB和正常视网膜组织进行差异基因表达(differentially expressed genes, DEGs)筛选,对其进行生物学功能、信号通路及差异基因的蛋白互作网络(protein-protein interaction, PPI)分析,从而帮助我们进一步研究和更好地探索RB发生发展的潜在机制。

1 材料与方法

1.1 基因芯片数据获取

从美国国立生物技术信息中心(National Center of Biotechnology Information, NCBI)公共数据平台GEO中获取2组基因芯片数据谱GSE97508芯片,平台是GPL15207 [(PrimeView) Affymetrix Human

Gene Expression Array)和GSE110811^[4],芯片平台是GPL16686 [(HuGene-2_0-st) Affymetrix Human Gene 2.0 ST Array transcript (gene) version],该2个芯片数据分别包括6个RB组织和3个正常视网膜组织和28个RB组织和3个正常视网膜组织。

1.2 筛选DEGs

采用GEO2R在线工具对2组芯片的RB组织标本与正常视网膜组织标本之间的DEGs进行筛选,筛选条件为 $|\log(\text{差异倍数FC})|>2$,差异显著性校正 $P<0.05$, $\log FC<0$ 为下调DEGs,而 $\log FC>0$ 为上调DEGs。然后对2组芯片筛选出的DEGs用在线Draw Venn Diagrams(<http://bioinformatics.psb.ugent.be/webtools/Venn/>)软件取交集进一步找出共同的DEGs。

1.3 GO注释及KEGG富集通路分析

基因本体注释(gene ontology analysis, GO)是一种常用的定义基因及其RNA或蛋白产物的方法,用于识别高通量转录组或基因组数据的独特生物学特性^[5]。GO注释目前包括3个方面的生物学内容:分子功能(molecular function, MF)、细胞成分(cellular components, CC)、生物过程(biological process, BP)。KEGG是1个关于基因组、疾病、生物途径、药物和化学物质的数据库集合^[6]。本研究采用在线STRING检索工具(search tool for the retrieval of interacting genes, STRING)对DEGs进行GO注释和KEGG富集通路分析,并对KEGG通路进行可视化图形展示($P<0.05$)。

1.4 PPI的可视化

采用STRING数据库对DEGs的PPI进行可视化^[7], 去除网络未连接点, 并进一步筛选出中心性最高, 即与其他节点关系最密切的核心基因。

2 结果

2.1 RB组织GDEs的识别

本研究中有34个RB组织和6个正常视网膜组织。通过GEO2R在线工具, 我们分别从基因芯片GSE97508和GSE110811中根据筛选条件为 $|\log(\text{差异倍数FC})| > 2$, 差异显著性校正 $P < 0.05$, 分别筛选出1 356和1 150个DEGs。然后利用Draw Venn Diagrams软件对2组DEGs进一步取交集共筛选到20个常见DEGs, 其中RB组织中有17个下调DEGs ($\log\text{FC} < 0$), 3个上调DEGs ($\log\text{FC} > 0$) (表1, 图1)。

2.2 RB的DEGs的GO注释和KEEG富集通路

20个RB的DEGs的GO分析结果如下, 1)上调DEGs对于生物过程(biological process, BP): DNA构象改变、染色体浓缩、调控有丝分裂细胞周期

过程、姐妹染色单体分离、核分裂、细胞分裂; 2)上调DEGs对于分子功能(molecular function, MF): DNA催化剂; 3)下调DEGs对于BP: 调节视紫红质介导的信号通路、光刺激的感觉、光感受器细胞的修复、视觉感知、磷酸盐代谢过程负调控、蛋白质修饰过程的负调控、类维生素A代谢过程; 4)下调DEGs对于MF: 细胞内环状核苷酸激活了阳离子通道活性、cGMP连接; 5)下调DEGs对于CC: 光感受器外节、睫状体、睫状体膜、细胞膜、质膜、感光细胞内外节、神经投射、高尔基体膜和突触(表2)。KEGG通路分析结果表明: DEGs富集于光传导信号通路, 其中包括CNGA1, CNGB1, RHO, SAG四个基因均为下调DEGs, 并对通路进行可视化图形展现(表3, 图2)。而上调DEGs没有显著的信号通路($P < 0.05$)。

2.3 PPI

使用String在线工具对20个DEGs进行蛋白质与蛋白质相互作用网络的构建, 结果显示20节点和9条相互作用关系(图3)。发现与其他节点连接最紧密核心基因RHO, 连接基因包括CNGA1, CNGB1, SAG, CABP5和BCO2。

表1 20个DEGs

Table 1 20 DEGs

DEGs	基因名
上调基因	NUSAP1, TOP2A, MCM6
下调基因	BBS9, BCO2, SAG, CABP5, LRRC4C, RASSF2, SPRY2, CNGA1, CLUL1, UNC13C, FABP12, PCDH15, FBXO32, LRRC39, YPEL2, RHO, CNGB1

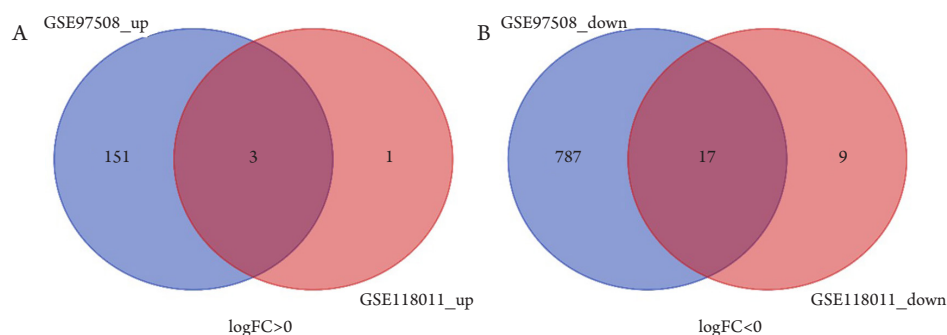


图1 通过Venn Diagrams软件对2组芯片数据(GSE97508和GSE118011)DEGs取交集共20个DEGs

Figure 1 Venn Diagrams are used to intersect two groups of chip data (GSE97508 and GSE118011) DEGs, and there are 20 DEGs in total

(A) 3个上调DEGs ($\log\text{FC} > 0$); (B) 17个下调DEGs ($\log\text{FC} < 0$).

(A) 3 up-regulated DEGS ($\log\text{FC} > 0$); (B) 17 decrement DEGs ($\log\text{FC} < 0$).

表2 DEGs的GO功能注释

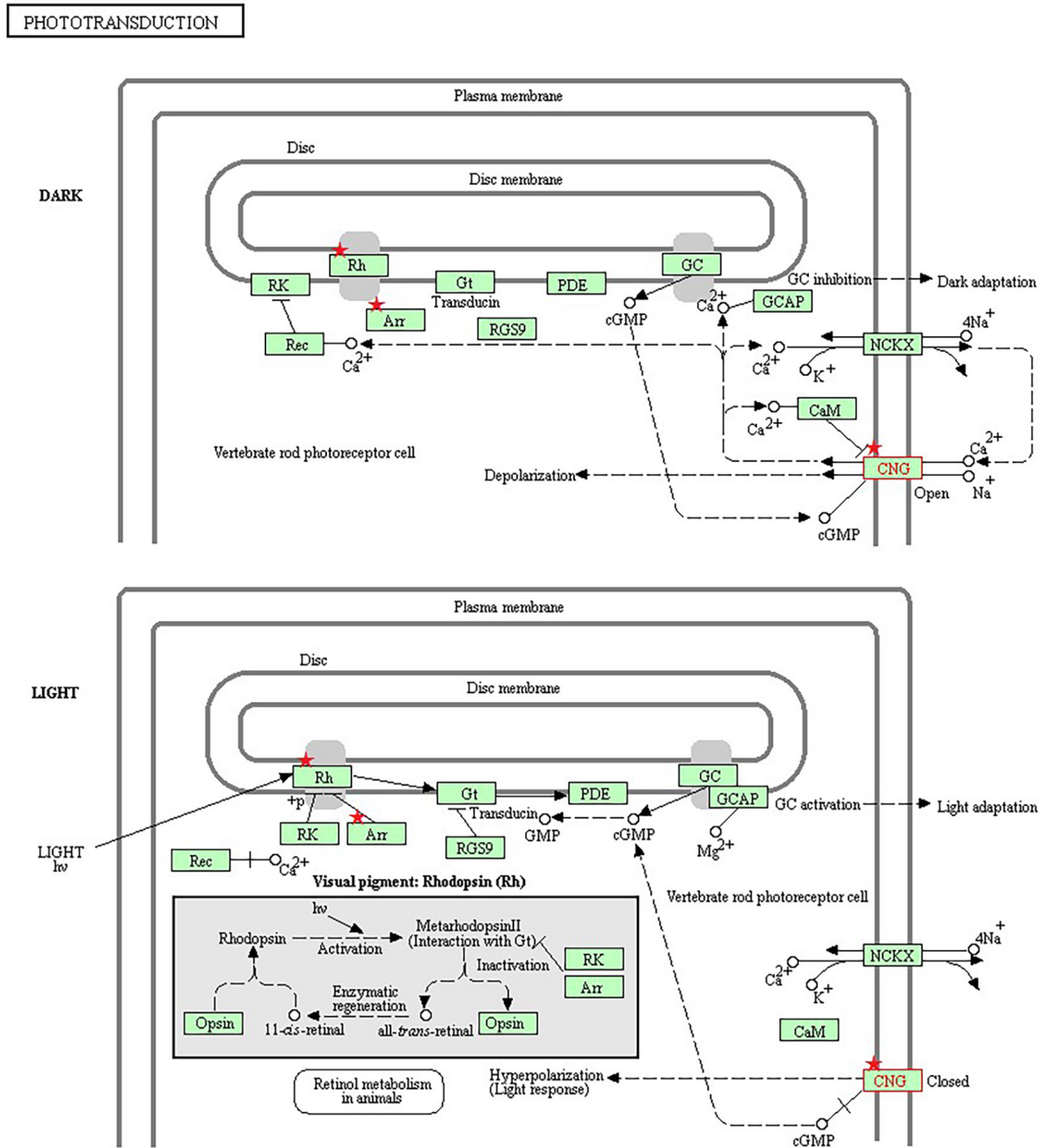
Table 2 GO analysis of DEGs

DEGs	类型	GO功能项目	基因数	百分比/%	FDR
上调	MF	GO:0140097~catalytic activity, acting on DNA	2	1.46	0.0135
	BP	GO:0071103~DNA conformation change	3	1.17	0.00047
	BP	GO:0030261~chromosome condensation	2	6.67	0.00078
	BP	GO:1903047~mitotic cell cycle process	3	0.53	0.0016
	BP	GO:0000819~sister chromatid segregation	2	1.63	0.0041
	BP	GO:0000280~nuclear division	2	0.75	0.0095
	BP	GO:0051301~cell division	2	0.41	0.0261
	BP	GO:0007346~regulation of mitotic cell cycle	2	0.33	0.0384
下调	MF	GO:0005221~intracellular cyclic nucleotide activated cation channel activity	2	22.22	0.004
	MF	GO:0030553~cGMP binding	2	15.38	0.004
	BP	GO:0016056~rhodopsin mediated signaling pathway	4	30.77	4.89E-07
	BP	GO:0022400~regulation of rhodopsin mediated signaling pathway	4	16.67	1.40E-06
	BP	GO:0050953~sensory perception of light stimulus	2	2.34	7.06E-05
	BP	GO:0045494~photoreceptor cell maintenance	3	8.33	0.00024
	BP	GO:0007601~visual perception	4	1.90	0.001
	BP	GO:0007600~sensory perception	6	0.67	0.0026
	BP	GO:0045936~negative regulation of phosphate metabolic process	4	0.71	0.0282
	BP	GO:0031400~negative regulation of protein modification process	4	0.68	0.0314
	BP	GO:0001523~retinoid metabolic process	2	2.30	0.0488
	CC	GO:0001750~photoreceptor outer segment	5	6.10	1.27E-06
	CC	GO:0044441~ciliary part	6	1.42	2.52E-05
	CC	GO:0060170~ciliary membrane	4	4.71	2.52E-05
	CC	GO:0031253~cell projection membrane	5	1.58	0.00012
	CC	GO:0120038~plasma membrane bounded cell projection part	8	0.61	0.00012
	CC	GO:0098590~plasma membrane region	7	0.66	0.00021
	CC	GO:0042622~photoreceptor outer segment membrane	2	16.67	0.00074
	CC	GO:0043005~neuron projection	6	0.53	0.0026
	CC	GO:0043195~terminal bouton	2	4.55	0.0062
	CC	GO:0001917~photoreceptor inner segment	2	4.35	0.0064
	CC	GO:0044459~plasma membrane part	8	0.30	0.0068
	CC	GO:0030660~Golgi-associated vesicle membrane	2	1.96	0.252
	CC	GO:0045202~synapse	4	0.47	0.0326

表3 DEGs的KEGG富集通路

Table 3 KEGG pathway of DEGs

通路ID	名称	基因数	占比/%	FDR	基因
hsa04744	Phototransduction	4	15.38	2.77E-07	CNGA1, CNGB1, RHO, SAG



04744 11/9/15
(c) Kanehisa Laboratories

图 2 DEGs 的 KEGG 富集在光传导通路, 标红色五角星为 4 个 DEGs, 其中 CNG 为 CNGA1, CNGB1, RH 为 RHO, Arr 为 SAG
Figure 2 KEGG of DEGs was enriched in the photoconductive pathway, and there were four DEGs marked with red pentagrams, among which CNG was CNGA1, CNGB1, RH was Rho, and Arr was SAG

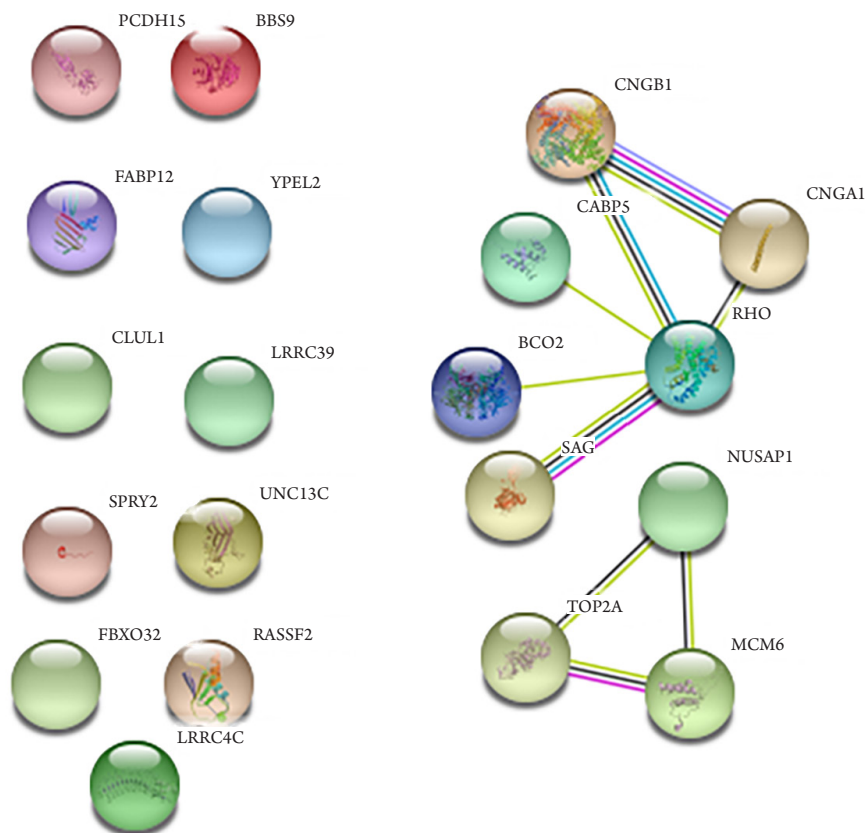


图3 使用String (<https://string-db.org/>) 在线工具对20个DEGs进行蛋白质与蛋白质相互作用网络的构建, 结果显示20节点和9条相互作用关系。与其他节点连接最紧密核心基因为RHO, 连接基因包括CNGA1, CNGB1, SAG, CABP5和BCO2

Figure 3 The String (<https://string-db.org/>) online tool showed PPI network of 20 DEGs, and the results showed 20 nodes and 9 interaction relationships. The most closely linked core gene was Rho, which included CNGA1, CNGB1, SAG, CABP5 and BCO2

3 讨论

RB是儿童最常见的眼内恶性肿瘤, 在RB早期及时诊断和干预, 患儿的存活率超过95%, 而一旦发生眼外转移, 其存活率低于50%^[8]。近年来对RB基因组学研究增多, 本研究利用生物信息学分析方法对GEO数据库下载的RB组织和正常视网膜组织的基因片数据进行DEGs分析, 并对其进行分类分析、功能富集分析和相互作用网络分析, 从而深入了解RB发生、发展的分子机制, 筛选可作为除RB基因以外的诊断的关键基因和治疗药物靶点。

在本研究通过RB组织和正常视网膜组织的比较, 共筛出20个DEGs, 其中上调基因3个, 下调基因17个, 应用在线STRING数据库进行GO功能注释结果显示表达上调基因所富集的功能主要

在细胞分裂、染色体浓缩、核分裂和DNA构象改变。下调基因主要富集在光传导、视觉感知、光感受器细胞修复、感光细胞内外节和调节视紫红质介导的信号通路等。这些生物学过程的调节异常是参与RB发生发展。KEGG通路显示上调基因没有显著的信号通路, 下调基因参与光传导信号通路, 其中包括CNGA1, CNGB1, RHO, SAG四个基因, 通过PPI网络提示这四个基因相互联系, 并发现与其他节点连接最紧密核心基因RHO。基因RHO(rhodopsin)为视紫红质, 早先Donoso等^[9]使用单克隆抗体组织病理研究在人胎儿视网膜和五个视网膜母细胞瘤中鉴定了视紫红质, 视网膜母细胞瘤中视紫红质和s抗原的存在可能有助于理解光感受器细胞的发育, 并参与视觉光转导蛋白的表达。包括与视觉直接相关的事件。之后Hurwitz等^[10]研究发现: 只有2种抗视紫红

质抗体识别肿瘤上的表位,而在正常人视网膜上完整识别的五种表位的视紫红质分子可能在视网膜母细胞瘤中不表达。二者均与本研究筛选出核心基因RHO为下调基因情况相符,因此视紫红质低表达为RB的发生发展提供了一个新的思路,由于近年鲜有对RHO与RB相关性研究,需要对其进一步实验验证,希望有助于成为新的抗肿瘤治疗靶点。本研究KEGG通路富集在光传导信号通路与Hurwitz等^[10]在RB中发现存在锥细胞特异性光转导级联反应相符,该研究推测该肿瘤沿着锥细胞谱系进行了生化分化。也与数据谱GSE11081发现感光细胞表达与视网膜母细胞瘤严重程度呈负相关有关联。因此考虑RB与正常视网膜组织差异基因通路可能与光信号转导有关。

本研究也有局限性,一是GSE11081对其提供28例RB组织进行间变质轻中重的分级,从而得出感光细胞表达与视网膜母细胞瘤严重程度呈负相关,而本研究并未对RB分级而是直接将RB组织与正常视网膜差异进行筛选,因此GSE11081研究表明的核孔蛋白表达增加,光感受器表达减少,是严重间变性RB的特征在本研究中并未发现。今后若有更多对RB轻中重分级的数据谱,可对其纳入及进行分级研究;二是基因芯片数据谱才2组偏少,且发现共同差异基因过少,这些对基因功能富集和通路分析意义都是比较有限的,期待今后能从更多的数据库纳入发现更多研究RB基因的数据谱,从而丰富差异基因。

综上所述,本文采用生物信息学方法对RB基因芯片数据进行挖掘,通过对RB发生发展的生物过程、分子功能、细胞定位和信号通路等分析,以期往能为RB发生的机制研究、肿瘤标志物的筛

选及药物靶点提供参考。

参考文献

1. Dimaras H, Corson TW. Retinoblastoma, the visible CNS tumor: A review[J]. *J Neurosci Res*, 2019, 97(1): 29-44.
2. Zhang J, Benavent AC, McEcoy J, et al. A novel retinoblastoma therapy from genomic and epigenetic analyses[J]. *Nature*, 2012, 481(7381): 329-334.
3. Bert V, Nickolas P, Velculescu VE, et al. Cancer genome landscapes[J]. *Science*, 2013, 339(6127): 1546-1558.
4. 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.
5. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium[J]. *Nat Genet*, 2000, 25(1): 25-29.
6. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes[J]. *Nucleic Acids Res*, 2000, 28(1): 27-30.
7. Szklarczyk D, Franceschini A, Wyder S, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life[J]. *Nucleic Acids Res*, 2015, 43: D447-D452.
8. Sameh E, Soliman HR, Cheng Y, et al. Genetics and molecular diagnostics in retinoblastoma—an update[J]. *Asia Pac J Ophthalmol*, 2017, 6(2): 197-207.
9. Donoso LA, Hamm H, Dietzschold B, et al. Rhodopsin and Retinoblastoma: a monoclonal antibody histopathologic study[J]. *Arch Ophthalmol*, 1986, 104(1): 111-113.
10. Hurwitz RL, Bogenmann E, Font RL, et al. Expression of the functional cone phototransduction cascade in retinoblastoma[J]. *J Clin Invest*, 1990, 85(6): 1872-1878.

本文引用: 张艳, 吴瑜瑜. 视网膜母细胞瘤基因表达谱芯片的生物信息学分析[J]. 眼科学报, 2021, 36(5): 326-332. doi: 10.3978/j.issn.1000-4432.2021.02.02

Cite this article as: ZHANG Yan, WU Yuyu. Bioinformatics analysis of differentially expressed genes in retinoblastoma[J]. *Yan Ke Xue Bao*, 2021, 36(5): 326-332. doi: 10.3978/j.issn.1000-4432.2021.02.02