

# 基于加权基因共表达网络识别糖尿病视网膜病变与免疫相关的关键基因

袁琳慧<sup>1,2</sup>, 张立军<sup>1,2</sup>, 刘新<sup>1,2</sup>, 祁媛媛<sup>1,2</sup>

引用:袁琳慧,张立军,刘新,等. 基于加权基因共表达网络识别糖尿病视网膜病变与免疫相关的关键基因. 国际眼科杂志 2023;23(8):1343-1351

作者单位:<sup>1</sup>(116033)中国辽宁省大连市,大连医科大学附属大连市第三人民医院眼科;<sup>2</sup>(116044)中国辽宁省大连市,大连医科大学

作者简介:袁琳慧,硕士,主治医师,研究方向:眼底病。

通讯作者:张立军,博士,主任医师,研究方向:角膜病、屈光手术。Lijunzhangw1970@163.com

收稿日期:2023-01-09 修回日期:2023-07-05

## 摘要

目的:通过生物信息学的方法探究糖尿病视网膜病变(DR)中免疫相关的关键基因以及免疫细胞的浸润情况。

方法:2022-09/10从GEO数据库获取基因芯片数据集,采用“limma”R包获得差异表达基因(DEGs),并进行GO功能注释和KEGG通路富集分析,基于CIBERSORT算法分析免疫细胞浸润情况。通过加权基因共表达网络分析(WGCNA)筛选与免疫相关基因模块中的DEGs,利用STRING在线数据库及Cytoscape软件构建蛋白质互作网络,利用MCODE以及cytoHubba插件进一步并筛选出关键基因。

结果:共筛选出上调差异基因1426个,下调差异基因206个。原始B细胞、浆细胞、记忆型CD4<sup>+</sup>T细胞、调节性T细胞(Tregs)、M0型巨噬细胞、M1型巨噬细胞以及中性粒细胞7种免疫细胞显著高表达( $P < 0.05$ );NK cells activated这种免疫细胞低表达( $P < 0.05$ )。WGCNA分析后,与免疫最相关模块中差异基因820个,构建PPI网络后利用插件筛选出10个关键基因,利用各差异基因在PPI中的相关性程度进一步筛选出2个关键基因为DLGAP5与AURKB。

结论:利用生物信息学的方式筛选出DR患者中免疫细胞浸润情况以及与免疫相关的关键基因,可为DR的进一步研究与诊疗提供依据。

关键词:糖尿病视网膜病变;生物信息学;加权共表达;免疫浸润;蛋白质互作网络;差异基因;免疫细胞

DOI:10.3980/j.issn.1672-5123.2023.8.20

## Identification of key immune related genes in diabetes retinopathy based on weighted gene co-expression network

Lin-Hui Yuan<sup>1,2</sup>, Li-Jun Zhang<sup>1,2</sup>, Xin Liu<sup>1,2</sup>, Yuan-Yuan Qi<sup>1,2</sup>

<sup>1</sup>Department of Ophthalmology, the Third People's Hospital Affiliated to Dalian Medical University, Dalian 116033, Liaoning

Province, China; <sup>2</sup>Dalian Medical University, Dalian 116044, Liaoning Province, China

Correspondence to: Li-Jun Zhang. Department of Ophthalmology, the Third People's Hospital Affiliated to Dalian Medical University, Dalian 116033, Liaoning Province, China; Dalian Medical University, Dalian 116044, Liaoning Province, China. Lijunzhangw1970@163.com

Received:2023-01-09 Accepted:2023-07-05

## Abstract

• AIM: To explore the key genes related to immunity and immune cell infiltration levels in diabetes retinopathy (DR) using bioinformatics.

• METHODS: Differential expression genes (DEGs) were obtained by “limma” R from Gene Expression Omnibus (GEO) data from September to October 2022, Gene ontology (GO) and Kyoto encyclopedia of genes and genomes (KEGG) were analyzed, and the infiltration of immune cell types in each sample was calculated based on CIBERSORT algorithm. Weighted gene co-expression network analysis (WGCNA) was used to screen for DEGs in immune-related gene modules. The protein-protein interaction (PPI) network was established by STRING online database and Cytoscape, and the hub genes were screened by MCODE and cytoHubba plug-ins.

• RESULTS: The results showed that 1426 up-regulated and 206 down-regulated differential genes were screened, where 7 immune cell types, including B cell naive, Plasma cells, CD4<sup>+</sup>T cells, T cells regulatory (Tregs), Macrophages M0, Macrophages M1 and Neutrophils were significantly overexpressed ( $P < 0.05$ ), while others were low expressed ( $P < 0.05$ ). After WGCNA, a total of 820 DEGs were found in the modules most related to immunity. After constructing the PPI network, 10 key genes were screened using plug-ins, and two key genes were further screened using the expression amount of each differential gene in PPI: DLGAP5 and AURKB.

• CONCLUSION: This study used bioinformatics to screen the infiltration of immune cells and key genes related to immunity in patients with DR. These findings may provide evidences for future research, diagnosis, and treatment of DR.

• KEYWORDS: diabetes retinopathy; bioinformatics; weighted gene co-expression network analysis; immune infiltration; protein-protein interaction network; differential gene; immune cell

Citation: Yuan LH, Zhang LJ, Liu X, et al. Identification of key immune related genes in diabetes retinopathy based on weighted gene co-expression network. *Guoji Yanke Zazhi (Int Eye Sci)* 2023; 23(8):1343-1351

## 0 引言

糖尿病视网膜病变(diabetic retinopathy, DR)是我国工作年龄人群中发病率最高的视网膜血管疾病<sup>[1]</sup>,其处理复杂而难治,而且周期长。DR是近年来国内外眼科临床治疗中较为棘手的疾病之一。为探寻更好的预防、诊断与治疗方式,关于DR的发病机制一直是研究热点。因此,我们仍需要进一步探究DR的遗传标记物和治疗靶点,为DR的诊断和干预提供新的方法。

视网膜是一种免疫特权组织。血-视网膜屏障(blood-retinal barriers, BRB)和免疫抑制保护视网膜免受外部和内部入侵和攻击<sup>[2-3]</sup>。在DR发生时,由于广泛的BRB损伤,视网膜免疫特权被破坏,免疫耐受性大大降低,大量循环免疫细胞浸润视网膜并激活先天或后天免疫反应,导致视网膜血管和神经元的进行性改变<sup>[4]</sup>。小胶质细胞是中枢神经系统(central nervous system, CNS)中先天免疫系统的吞噬细胞。视网膜和视神经是中枢神经系统的外延,小胶质细胞在眼部疾病的免疫发病机制中起着重要作用。小胶质细胞通过吞噬视网膜视神经节细胞(retinal ganglion cell, RGC)、分泌细胞毒性物质和破坏BRB的结构和功能,参与DR的发病机制<sup>[5]</sup>。DR的视网膜血管损伤也与补体系统的激活密切相关<sup>[6]</sup>。高糖环境刺激局部和全身炎症因子和趋化因子的表达增加,由此导致持续存在的炎症反应被认为会促进DR中血管和免疫系统的损伤,诱导BRB的破坏,进而导致黄斑水肿和视网膜新生血管的产生<sup>[7]</sup>。上述研究表明,免疫在DR中起着重要作用。因此,本研究旨在研究DR中免疫细胞的浸润情况以及寻找DR中与免疫相关差异基因,为DR的诊断与治疗提供新的思路。

## 1 对象和方法

**1.1 对象** 本研究2022-09/10使用GEO数据集为GSE179568, GPL15433平台,更新于2021-09-02<sup>[1]</sup>,下载于GEO网站(<http://www.ncbi.nlm.nih.gov/geo/>)。该数据集包括7个DR患者的视网膜前纤维血管增殖膜,7个特发性黄斑裂孔患者的视网膜内界膜以及10个特发性黄斑前膜患者的黄斑前膜的基因表达量芯片检测结果为对照组。

## 1.2 方法

**1.2.1 筛选差异表达基因** 在R语言环境下,利用“limma”包对数据集中DR样本以及对照样本的差异表达基因进行鉴定,有差异的判定标准为校正后的 $P < 0.05$ 且 $\log_2(\text{FoldChange})$ 绝对值 $> 1$ <sup>[8]</sup>。筛选出DR患者的差异表达基因(differentially expressed genes, DEGs)。

**1.2.2 GO功能注释和KEGG信号通路富集分析** 将筛选出来的差异表达基因导入DAVID 6.8数据库依据GO(Gene Ontology)数据库以人源基因为背景对差异表达基因进行生物学功能注释并利用KEGG(kyoto encyclopedia of genes and genomes)通路数据库进行信号通路的富集寻找DEGs所富集的信号通路<sup>[9]</sup>。

**1.2.3 免疫浸润CIBERSORT分析** CIBERSORT基于基因表达数据对免疫相关通路进行分析,从而对特定免疫细胞浸润情况进行考察<sup>[10]</sup>。本研究采用CIBERSORT方法分析DR患者与对照组间差异的免疫细胞表达水平。

**1.2.4 加权基因共表达分析** 使用加权基因共表达分析(weighted gene co-expression network analysis, WGCNA)分析挑选出与免疫通路相关的共表达编码基因模块。首先对样本进行聚类,剔除样本中的离群众后,对样本进行聚类。筛选出最佳软阈值 $\beta$ ,鉴定出共表达模块,对各个模块与免疫细胞分数间的相关性进行分析<sup>[11]</sup>。选取免疫浸润分数差异最显著的模块。

**1.2.5 蛋白质互作网络与关键基因的筛选** 免疫浸润分数差异最显著的模块与DEGs取交集后的结果进行蛋白质互作网络(protein-protein interaction network, PPI network)分析,采用STRING在线数据库,去除没有相互作用的孤立点,得到蛋白相互作用网络<sup>[12]</sup>,将该网络导入Cytoscape软件,使用MCODE插件识别出总网络中相关性最强的子网络<sup>[13]</sup>,使用CytoHubba插件筛选出总网络中相关性最强的10个基因<sup>[14]</sup>。根据MCODE以及cytoHubba的结果交互选出关键(hub)基因。

**统计学分析:**筛选差异表达基因、GO功能注释和KEGG信号通路富集分析、免疫浸润CIBERSORT分析、加权基因共表达分析数据分析以及可视化采用R V.3.4.3软件进行。蛋白质互作网络与关键基因的筛选数据分析以及可视化采用STRING在线数据库进行。有差异的判定标准为校正后的 $P < 0.05$ 且 $\log_2(\text{FoldChange})$ 绝对值 $> 1$ 。

## 2 结果

**2.1 基因差异表达结果** 数据经过处理后,根据 $P < 0.05$ 且 $\log_2(\text{FoldChange})$ 绝对值 $> 1$ 的标准,DR患者与对照组相比,共有1632个基因差异表达显著,其中1426个基因表达上调,206个基因表达下调,见图1、2。

**2.2 GO和KEGG富集分析结果** 对差异表达的基因进行GO富集分析,并对前20个最密集的通路进行可视化(图3A)。对差异表达的基因进行KEGG富集分析,并对前20个最密集的通路进行可视化(图3B)。

**2.3 免疫浸润CIBERSORT分析结果** 在22个免疫细胞中,经过免疫浸润CIBERSORT分析,对各个样本免疫细胞占比可视化,见图4、5。DR患者和对照组免疫细胞分数

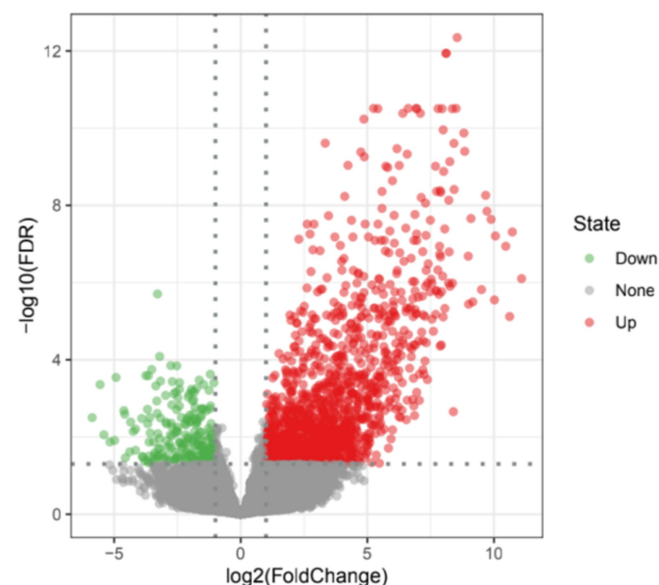


图1 基因差异表达分析火山图。

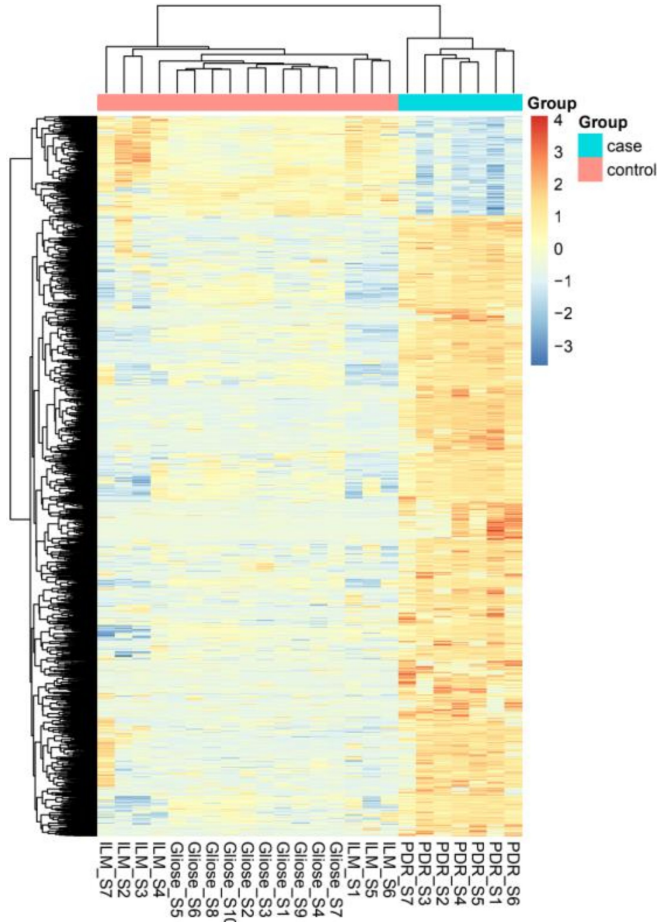


图2 基因差异表达热图。

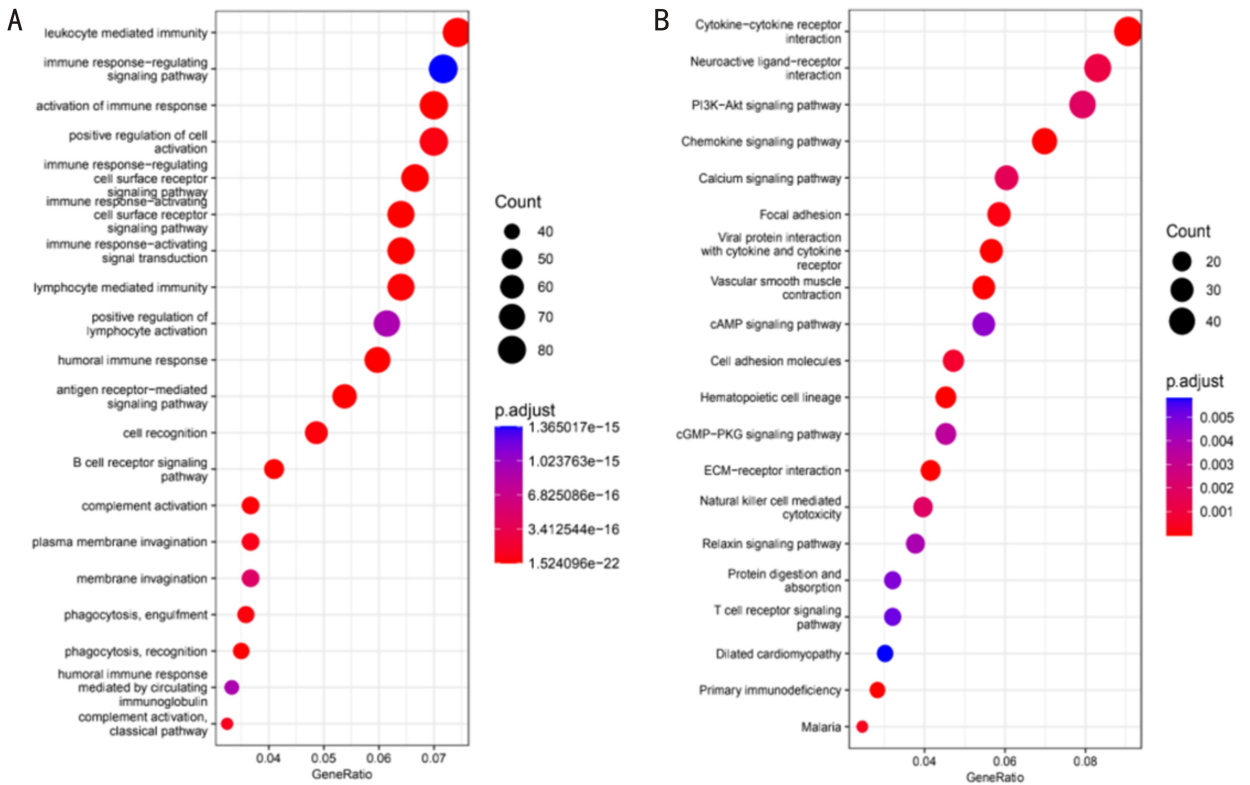


图3 GO和KEGG富集分析结果 A:GO富集;B:KEGG富集。

进行统计学检验,以对照组相比,DR样本中:原始B细胞、浆细胞、记忆型CD4<sup>+</sup>T细胞、调节性T细胞(Tregs)、M0型巨噬细胞、M1型巨噬细胞以及中性粒细胞7种免疫

细胞显著高表达,差异均有统计学意义( $P < 0.05$ );NK cells activated这种免疫细胞低表达,差异有统计学意义( $P < 0.05$ ),见图6。

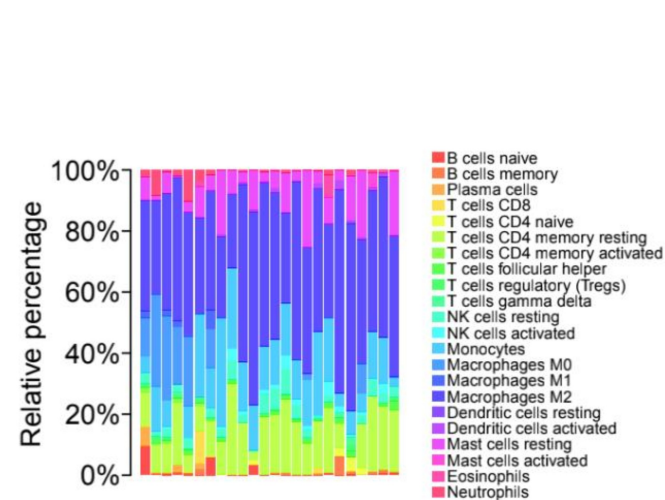


图4 CIBERSORT 堆叠图。

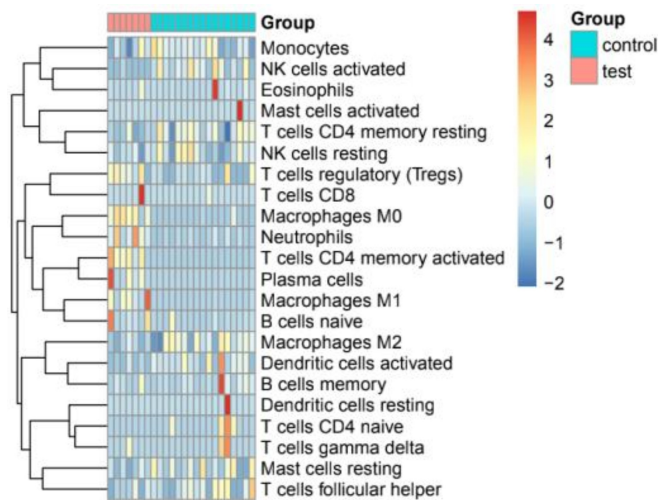


图5 免疫浸润差异热图。

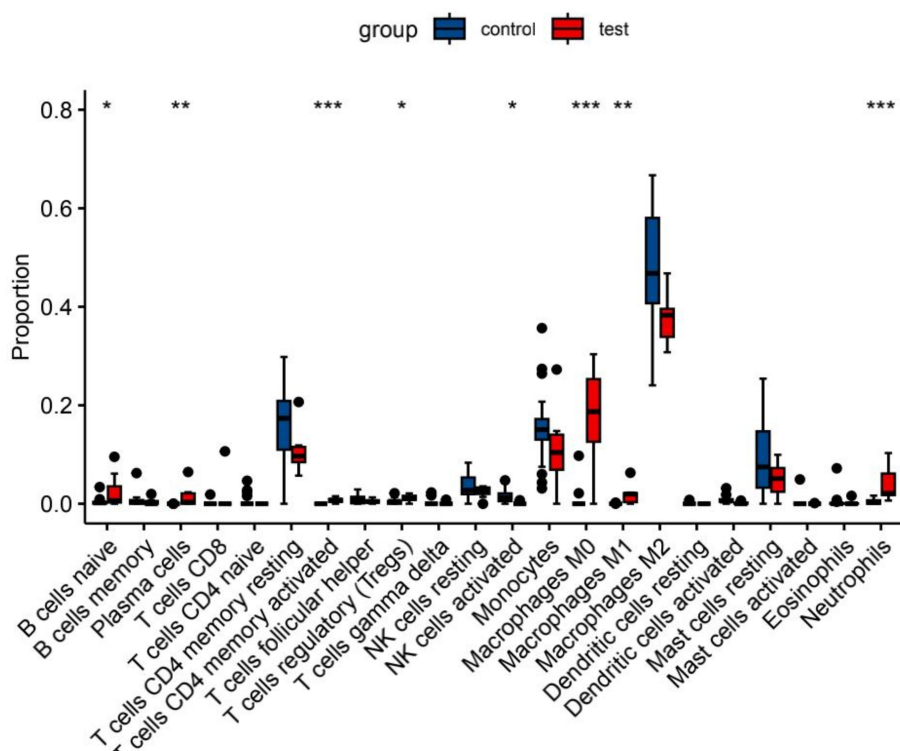


图6 CIBERSORT 差异箱线图 \* : $P < 0.05$ ; \*\* : $P < 0.01$ ; \*\*\* : $P < 0.001$ 。

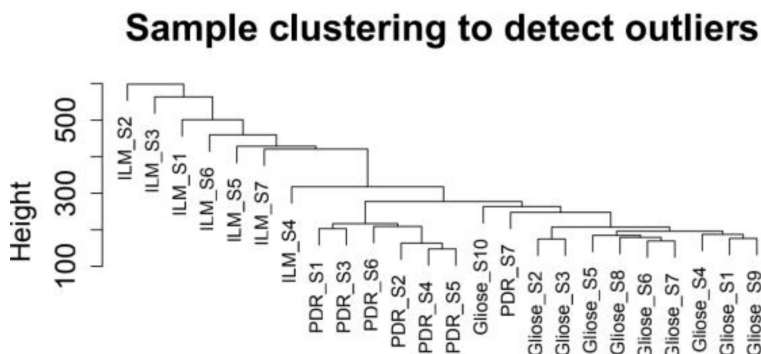


图7 样本聚类。

2.4 WGCNA 基因加权共表达结果 对样本聚类后,利用欧式距离进行层级聚类(图7),剔除离群值后进行样本聚类(图8)。根据算法选择筛选软阈值(图9),对共表达模

块进行鉴定并合并相似模块后得到26个共表达模块(图10)。各个模块与免疫细胞分数间的相关性进行分析并可视化(图11)。当 $P < 0.05$ 时,选取相关性系数最高

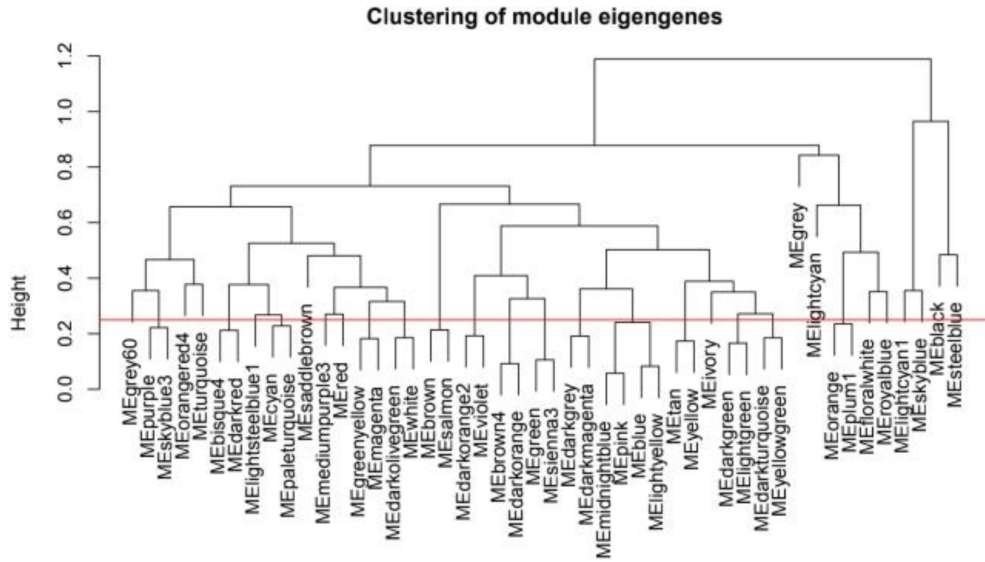


图 8 剔除离群值后样本聚类。

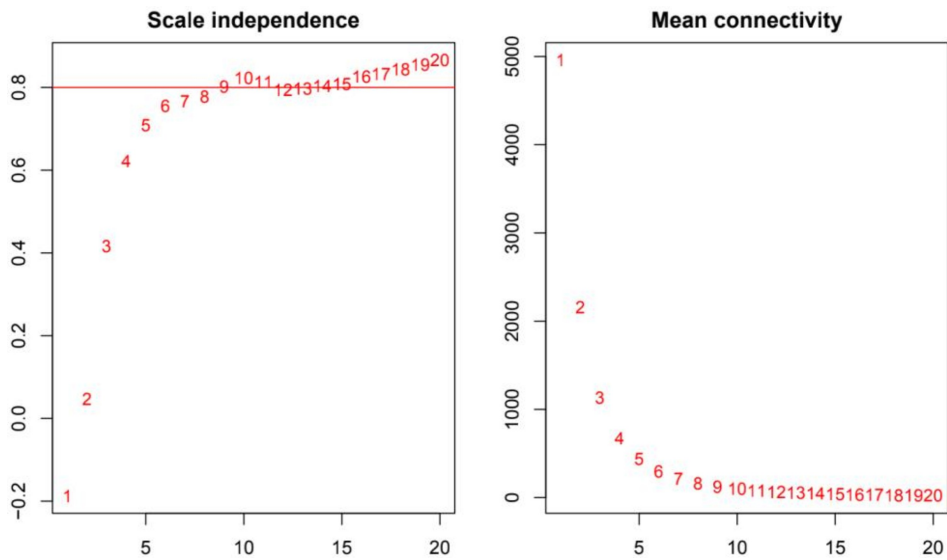


图 9 软阈值筛选。

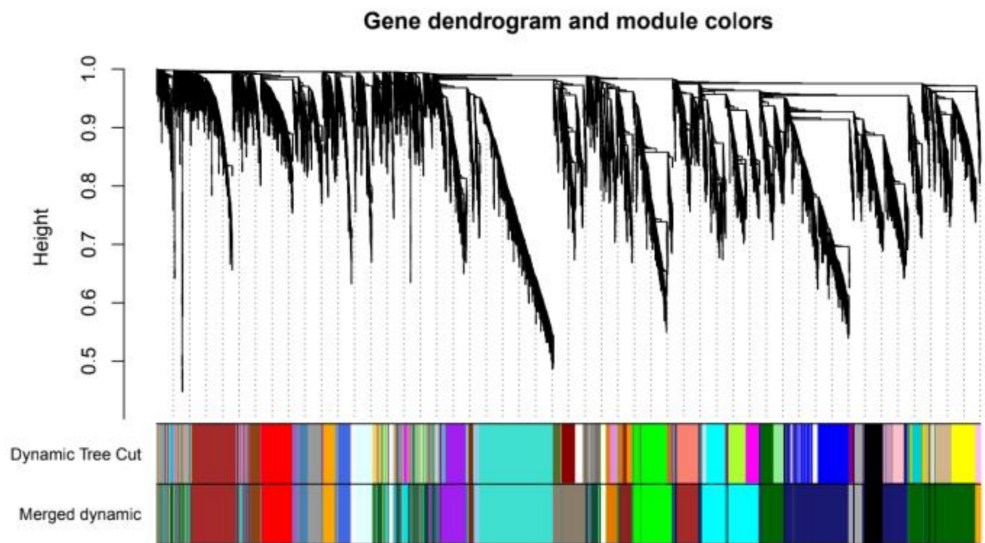


图 10 WGCNA 模块。

的模块为中心模块。此模块与 T cells CD4 memory activated 相关性系数最高达 0.84,因此选择此模块用于后续分析。此模块中的成员与基因显著性的相关性可视化呈现(图 12)。

2.5 PPI 与关键基因的筛选 将棕色模块中的基因与 2.1 得到的 DEGs 取交集,筛选出棕色模块中的差异基因 820 个。将这 820 个基因进行蛋白质互作分析,代入 STRING 数据库中,得到蛋白质互作网络(图 13)。利用 Cytoscape

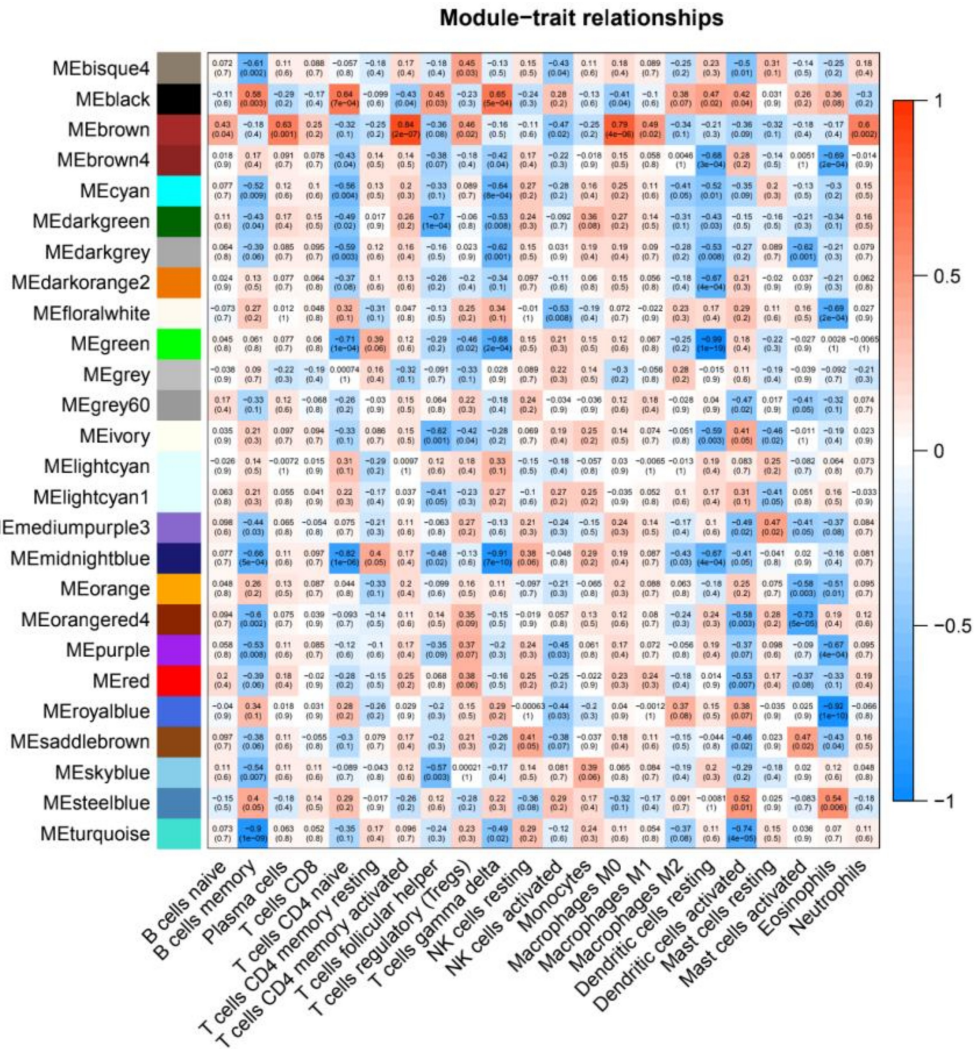


图 11 共表达模块与免疫细胞分数的相关性。

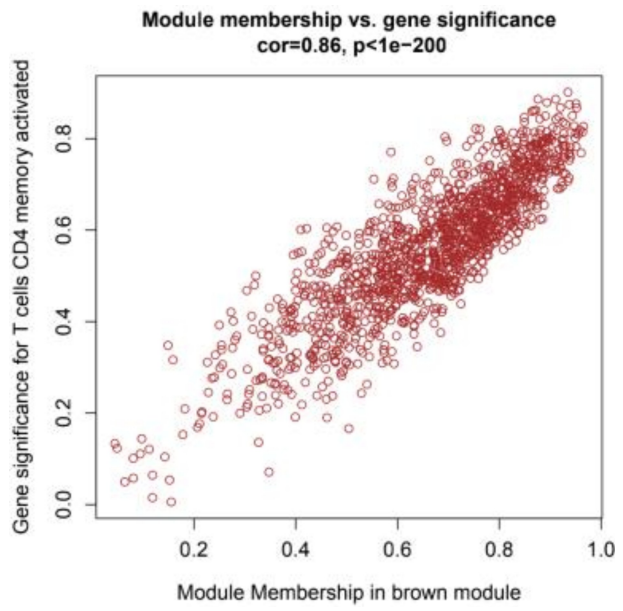


图 12 棕色模块与基因的显著性的相关性。

软件中的 MCODE 插件识别核心子网络 (图 14) 与 CytoHubba 插件得到的 top 表达基因相交得到 10 个基因: TOP2A、KIF15、CDC20、NCAPC、KIF4A、CEP55、AURKB、BUB1B、PBK、DLGAP5 (图 15)。根据这 10 个基因在 PPI 网

络中相关性程度,选取相关程度前 50 的为关键基因,最终筛选出 2 个关键基因:DLGAP5 以及 AURKB (图 16)。

**3 讨论**

本研究通过生物信息学的方法筛选 DR 患者视网膜组

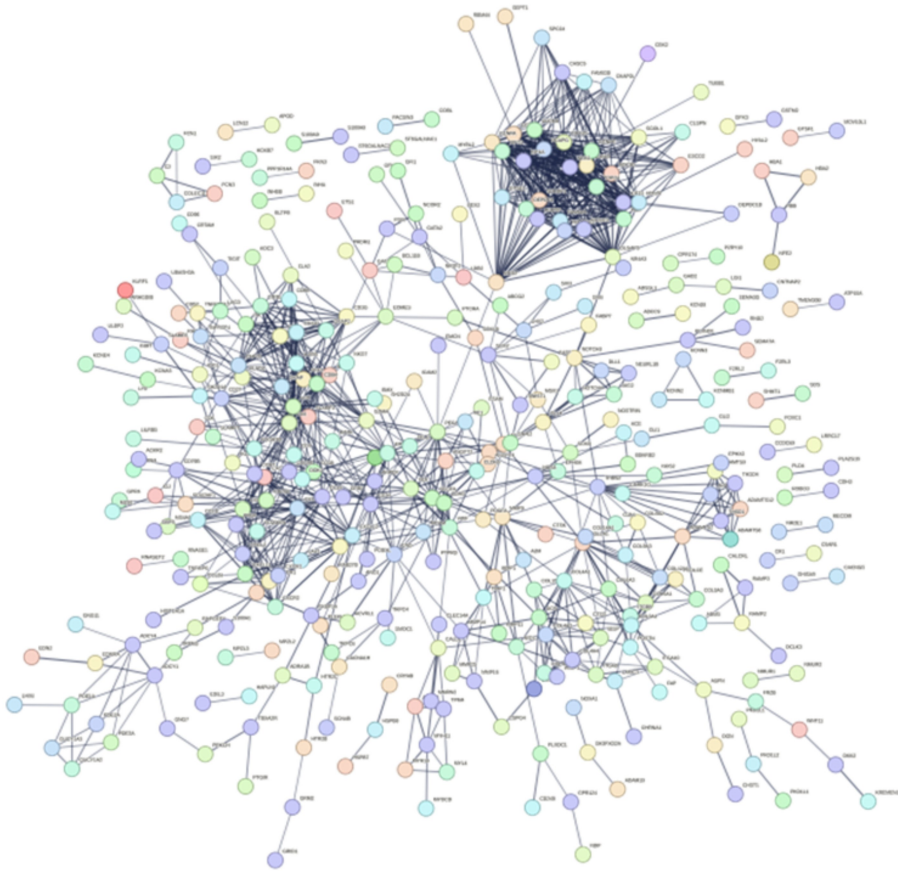


图 13 PPI 网络。

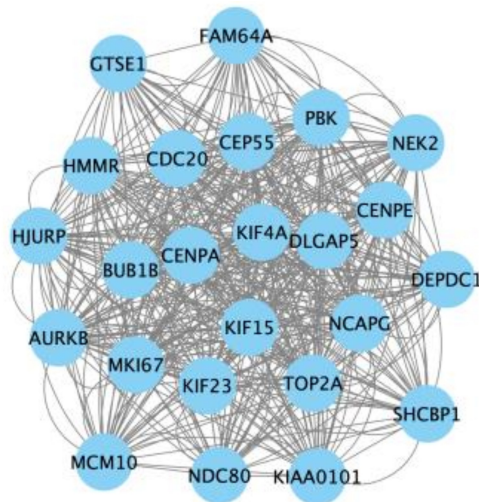


图 14 MCODE 子网络。

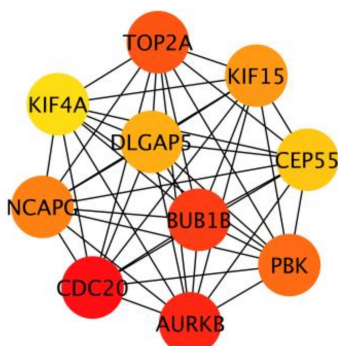


图 15 CytoHubba top 表达基因。

织与对照组的 DEGs,GO 富集结果发现差异基因均富集于与免疫反应相关的通路上,KEGG 富集于细胞因子受体作用通路、神经活性配体受体相互作用通路、PI3K-Akt 信号通路、化学信号通路以及钙信号通路等。

通过免疫浸润分析筛选出与 DR 高度相关的免疫细胞后结合 WGCNA 的方法再此筛选出免疫相关的 DEGs,利用 PPI 中的两组插件结合差异基因的表达量筛选出 2 个关键基因:DLGAP5 以及 AURKB。这 2 个关键基因在 DR 组中均表达上调。由于 DR 是慢性炎症性损伤导致的血管损伤性疾病<sup>[7]</sup>,所以炎症相关的差异基因上调比例远远高于下调,在生物信息学分析的过程中,不断缩小筛选

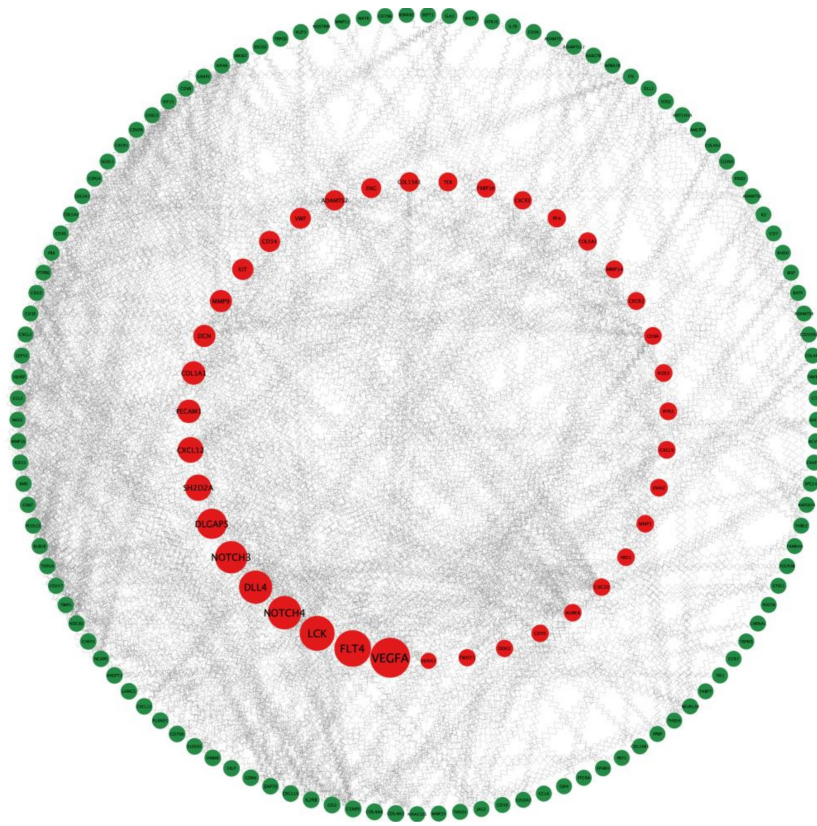


图 16 PPI 相关性可视化。

范围后,下调基因被过滤掉,导致本研究结果中的关键基因表达均呈现上调状态。

免疫浸润分析后,原始 B 细胞、浆细胞、记忆型 CD4<sup>+</sup>T 细胞、调节性 T 细胞 (Tregs)、M0 型巨噬细胞、M1 型巨噬细胞以及中性粒细胞 7 种免疫细胞显著高表达,其中活化的记忆型 CD4<sup>+</sup>T 细胞模块相关性最高且差异表达最为显著 ( $P < 0.001$ )。这一结果证明 DR 与免疫系统的激活密切相关。且研究表明,记忆型 CD4<sup>+</sup>T 细胞的活化主要依赖于葡萄糖的糖酵解过程<sup>[15]</sup>,在对于 1 型糖尿病的研究中发现,记忆型 CD4<sup>+</sup>T 细胞在 1 型糖尿病患者体内持续存在且高表达,这会导致难以通过免疫抑制治疗改善患者的胰岛素抵抗情况<sup>[16-18]</sup>。我们推测,记忆型 CD4<sup>+</sup>T 细胞的存在 DR 中同样会导致视网膜免疫抑制的破坏,促进 DR 的发生与发展。

Discs 大同源相关蛋白 5 (discs large homolog associated protein 5, DLGAP5) 以及极光激酶 B (AURKB) 均可影响细胞的有丝分裂过程,且均可在纺锤体上定位到。DLGAP5 是一种细胞周期调控基因的表达产物,被证实多种癌症疾病中通过调控细胞周期产生作用,随着 DLGAP5 表达量的升高,癌症细胞的侵袭性也呈现上升状态,所以其被认为是一种致癌蛋白<sup>[19-21]</sup>。DLGAP5 也被认为是潜在的 1 型糖尿病的靶点<sup>[22]</sup>。DLGAP5 可参与激活 NF- $\kappa$ B<sup>[23]</sup>, NF- $\kappa$ B 通路可通过各种生长因子、促血管因子促进 DR 的发展<sup>[24,25]</sup>。极光激酶 B (AURKB) 是极光激酶家族成员之一,为蛋白质丝氨酸/苏氨酸激酶。是一个重要的有丝分裂调节因子,参与多种有丝分裂事件的调节,同样, AURKB 也在多种癌症中起到促进癌症发生、发展的

作用<sup>[26-28]</sup>。在 DR 的相关研究中,实验证明在高氧诱导视网膜新生血管 (oxygen-induced retinopathy, OIR) 小鼠模型视网膜中 AURKB 表达增高,抑制 AURKB 活性可以抑制血管内皮细胞增殖、迁移和小管形进而抑制新生血管的产生<sup>[29]</sup>。

综上,本研究筛选了 DR 中 2 个免疫相关靶基因,所有这些基因都富集于记忆型 CD4<sup>+</sup>T 细胞相关通路。DLGAP5 与 AURKB 可作为 DR 诊断和治疗的潜在靶点。然而,仍需要进一步的研究来确认靶基因与免疫浸润之间的关系,并确定靶基因和免疫浸润谱在 DR 发生、发展中的作用。

#### 参考文献

- 1 Boneva SK, Wolf J, Hajdú RI, *et al.* In - Depth Molecular Characterization of Neovascular Membranes Suggests a Role for Hyalocyte - Myofibroblast Transdifferentiation in Proliferative Diabetic Retinopathy. *Front Immunol* 2021;12:757607
- 2 Chen M, Luo C, Zhao J, *et al.* Immune regulation in the aging retina. *Prog Retin Eye Res* 2019;69:159-172
- 3 Stepp MA, Menko AS. Immune responses to injury and their links to eye disease. *Transl Res* 2021;236:52-71
- 4 Xu H, Chen M. Diabetic retinopathy and dysregulated innate immunity. *Vision Res* 2017;139:39-46
- 5 Lim RR, Wieser ME, Ganga RR, *et al.* NOD-like Receptors in the Eye: Uncovering Its Role in Diabetic Retinopathy. *Int J Mol Sci* 2020;21(3):899
- 6 Xu H, Chen M. Targeting the complement system for the management of retinal inflammatory and degenerative diseases. *Eur J Pharmacol* 2016;787:94-104
- 7 Mesquida M, Drawnel F, Fauser S. The role of inflammation in diabetic eye disease. *Semin Immunopathol* 2019;41(4):427-445



- 8 Liu S, Wang Z, Zhu R, *et al.* Three Differential Expression Analysis Methods for RNA Sequencing: limma, EdgeR, DESeq2. *J Vis Exp* 2021;175
- 9 Zhang C, Zheng Y, Li X, *et al.* Genome-wide mutation profiling and related risk signature for prognosis of papillary renal cell carcinoma. *Ann Transl Med* 2019;7(18):427
- 10 Kawada JI, Takeuchi S, Imai H, *et al.* Immune cell infiltration landscapes in pediatric acute myocarditis analyzed by CIBERSORT. *J Cardiol* 2021;77(2):174-178
- 11 Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008;9:559
- 12 von Mering C, Huynen M, Jaeggi D, *et al.* STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res* 2003;31(1):258-261
- 13 Liu Y, Jin J, Chen Y, *et al.* Integrative analyses of biomarkers and pathways for adipose tissue after bariatric surgery. *Adipocyte* 2020;9(1):384-400
- 14 Zhou S, Lu H, Xiong M. Identifying Immune Cell Infiltration and Effective Diagnostic Biomarkers in Rheumatoid Arthritis by Bioinformatics Analysis. *Front Immunol* 2021;12:726747
- 15 刘悦越, 国泰, 李琦涵. 记忆性 CD4+T 细胞的形成及其影响因素. 细胞与分子免疫学杂志 2015;31(9):1267-1269,1275
- 16 Ehlers MR, Rigby MR. Targeting memory T cells in type 1 diabetes. *Curr Diab Rep* 2015;15(11):84
- 17 Fazeli P, Talepoor AG, Faghhi Z, *et al.* The frequency of CD4+ and CD8+ circulating T stem cell memory in type 1 diabetes. *Immun Inflamm Dis* 2022;10(10):e715
- 18 Eichmann M, Baptista R, Ellis RJ, *et al.* Costimulation Blockade Disrupts CD4+T Cell Memory Pathways and Uncouples Their Link to Decline in Cell Function in Type 1 Diabetes. *J Immunol* 2020;204(12):3129-3138
- 19 Zheng R, Shi Z, Li W, *et al.* Identification and prognostic value of DLGAP5 in endometrial cancer. *Peer J* 2020;8:e10433
- 20 Nie M, Wang Y, Yu Z, *et al.* AURKB promotes gastric cancer progression via activation of CCND1 expression. *Aging (Albany NY)*. 2020;12(2):1304-1321
- 21 Xiao J, Zhang Y. AURKB as a Promising Prognostic Biomarker in Hepatocellular Carcinoma. *Evol Bioinform Online* 2021;17:11769343211057589
- 22 Tang N, Dou X, You X, *et al.* Pan-cancer analysis of the oncogenic role of discs large homolog associated protein 5 (DLGAP5) in human tumors. *Cancer Cell Int* 2021;21(1):457
- 23 Branchi V, García SA, Radhakrishnan P, *et al.* Prognostic value of DLGAP5 in colorectal cancer. *Int J Colorectal Dis* 2019;34(8):1455-1465
- 24 Zheng Y, Liu L, Ye J. Identification of dysregulated modules based on network entropy in type 1 diabetes. *Exp Ther Med* 2018;15(4):3211-3214
- 25 魏丽娟, 苏秀兰. DLGAP5 的致癌机制以及在非小细胞肺癌诊断与治疗中的价值. 中国组织化学与细胞化学杂志 2020;29(1):87-91
- 26 Liang WJ, Yang HW, Liu HN, *et al.* HMGB1 upregulates NF- $\kappa$ B by inhibiting IKB- $\alpha$  and associates with diabetic retinopathy. *Life Sci* 2020;241:117146
- 27 Zhang T, Ouyang H, Mei X, *et al.* Erianin alleviates diabetic retinopathy by reducing retinal inflammation initiated by microglial cells via inhibiting hyperglycemia-mediated ERK1/2 - NF -  $\kappa$ B signaling pathway. *FASEB J* 2019;33(11):11776-11790
- 28 Borah NA, Reddy MM. Aurora Kinase B Inhibition: A Potential Therapeutic Strategy for Cancer. *Molecules* 2021;26(7):1981
- 29 宦晨阳. AURKB 在视网膜新生血管中的作用与机制研究. 中国人民解放军海军军医大学 2022