• 数据挖掘与循证医学 •

结合 Bulk RNA 和 Single-cell RNA 测序探索年龄相关性黄斑变性能量代谢关键基因及调控中药化合物预测

史随随[1]，罗越毅[2]，唐梦丹[3]，祁宝玉[3]，周　剑[3*]，巢国俊[2*]

1. 石河子大学医学院，新疆 石河子　832000
2. 中国中医科学院眼科医院，北京　100040
3. 北京中医药大学东方医院，北京　100078

摘　要：目的　探寻年龄相关性黄斑变性（age-related macular degeneration，AMD）能量代谢的关键基因和潜在调控的中药化合物。**方法**　从 GEO 数据库获取 AMD 相关 Bulk RNA 测序数据集 GSE135092，利用 R 语言筛选差异表达基因，并与 GSEA 数据库下载的能量代谢相关基因（energy metabolism-related genes，EMRGs）映射取交集，获取 AMD 差异表达 EMRGs，通过 Cytoscape 软件构建蛋白质相互作用（protein-protein interaction，PPI）、转录因子（transcription factor，TF）、微小 RNA（microRNA，miRNA）网络。使用 LASSO 回归在差异表达 EMRGs 中筛选 AMD 标志物，构建风险评分。同时，将 AMD 相关 Single-cell RNA 测序数据集 GSE188280 进行主成分分析（principal component analysis，PCA）及统一流形逼近与投影（uniform manifold approximation and projection，UMAP）分析以聚类降维，并进行细胞类型注释，进一步评估 AMD 关键差异表达 EMRGs 在不同细胞类型间的表达水平。将关键差异表达 EMRGs 导入 Coremine Medical 数据库中进行相关中药活性成分预测及分子对接、分子动力学验证。**结果**　筛选出 6 个 AMD 关键差异表达 EMRGs：*FOXO3*、*GSK3B*、*MEF2A*、*NCOA1*、*HDAC1*、*PPARG*，并构建相应 PPI、TF-mRNA、mRNA-miRNA 网络。LASSO 回归得出 *NCOA1*、*MEF2A*、*FOXO3*、*GSK3B* 表达与 AMD 发生呈正相关，*HDAC1* 表达与 AMD 发生呈负相关。分析 AMD 相关 Single-cell RNA 测序将 32 714 个细胞分类为 19 个簇，使用 marker 基因将细胞簇鉴定为 9 种细胞类型，进一步得到 *HDAC1*、*MEF2A*、*FOXO3* 及 *NCOA1* 在内皮细胞中高表达。根据关键基因 *HDAC1*、*MEF2A*、*FOXO3* 筛选到 9 味中药及其 285 种活性成分，经分子对接及分子动力学模拟，其中丹参醇 B 治疗 AMD 潜力最大。**结论**　*HDAC1* 和丹参醇 B 为 AMD 能量代谢的潜在关键基因及作用化合物。

关键词：年龄相关性黄斑变性；RNA 测序；能量代谢；关键基因；组蛋白脱乙酰酶 1；丹参醇 B

中图分类号：R285　　**文献标志码：**A　　**文章编号：**0253 - 2670(2024)07 - 2303 - 13

DOI: 10.7501/j.issn.0253-2670.2024.07.017

Exploration of key energy metabolism-related genes in age-related macular degeneration combined with Bulk RNA and Single-cell RNA sequencing and prediction of regulated traditional Chinese medicine compounds

SHI Suisui[1], LUO Yueyi[2], TANG Mengdan[3], QI Baoyu[3], ZHOU Jian[3], CHAO Guojun[2]

1. School of Medicine, Shihezi University, Shihezi 832000, China
2. Eye Hospital, China Academy of Chinese Medical Sciences, Beijing 100040, China
3. Dongfang Hospital, Beijing University of Chinese Medicine, Beijing 100078, China

Abstract: Objective　To explore the key energy metabolism-related genes (EMRGs) in age-related macular degeneration (AMD) and the potential regulated traditional Chinese medicine compounds. **Methods**　The AMD related Bulk RNA sequencing dataset GSE135092 was obtained from GEO, and the differentially expressed genes (DEGs) were screened by R software. Next, the intersection of DEGs with EMRGs downloaded from the GSEA was conducted to obtain AMD differentially expressed EMRGs, and

收稿日期：2023-09-30

基金项目：中国中医科学院科技创新项目（CI2021A02611）；北京市石景山区名医传承工作室师承项目（石卫健中医发[2020]6 号）；国家自然科学基金项目（8187150711）；国家中医临床研究基地业务建设科研专项课题项目（JDZX2015285）；石河子大学 2023 年自主支持科研项目（ZZZC2023024）

作者简介：史随随（1990—），女，博士，讲师，研究方向为中西医防治眼病。E-mail: 1043787258@qq.com

***通信作者：**周　剑（1964—），男，硕士，主任医师，研究方向为中西医防治眼病。E-mail: jian_zhou_dfyy@outlook.com

巢国俊（1965—），男，学士，主任医师，研究方向为中医防治眼病。E-mail: guojun_chao@outlook.com

protein-protein interaction (PPI), transcription factor (TF), and microRNA (miRNA) network were constructed using Cytoscape software. Then, LASSO regression was performed to screen AMD markers in differentially expressed EMRGs and construct a risk score. Meanwhile, the AMD related single-cell RNA sequencing dataset GSE188280 was used to conduct principal component analysis (PCA) and uniform manifold approximation and projection (UMAP) analysis for dimensionality reduction and cluster, and making cell type annotation, and further evaluating the expression levels of key differentially expressed EMRGs in AMD among different cell types. Finally, key differentially expressed EMRGs were imported into the Coremine Medical database for prediction of relevant traditional Chinese medicine active ingredients, molecular docking and molecular dynamics verification. **Results** A total of six key differentially expressed EMRGs were screened in AMD: *FOXO3*, *GSK3B*, *MEF2A*, *NCOA1*, *HDAC1* and *PPARG*. Corresponding PPI network, TF-mRNA and mRNA-miRNA regulatory networks were constructed. LASSO regression showed that the expression of *NCOA1*, *MEF2A*, *FOXO3*, and *GSK3B* was positively correlated with AMD, while the expression of *HDAC1* was negatively correlated with AMD. A total of 32 714 cells from AMD related single-cell RNA sequencing analysis were classified into 19 clusters, and the cell clusters were identified as nine cell types using marker genes. Furthermore, it was found that *HDAC1*, *MEF2A*, *FOXO3* and *NCOA1* were highly expressed in endohelial cells. Finally, a total of 285 active ingredients from nine traditional Chinese medicines were screened through key genes *HDAC1*, *MEF2A* and *FOXO3*. After molecular docking and molecular dynamics simulation, danshenol B was the most potential therapeutic compound in the treatment of AMD. **Conclusion** *HDAC1* and danshenol B are potential key energy metabolism-related gene and compound in AMD.

Key words: age-related macular degeneration; RNA sequencing; energy metabolism; key genes; HDAC1; danshenol B

年龄相关性黄斑变性（age-related macular degeneration，AMD）是与年龄相关的视网膜黄斑进行性退变疾病，造成患者不可逆性中心视力下降、视物变形，为世界卫生组织三大重点致盲眼病之一[1]。近 30 年中国 AMD 疾病负担居全球前列，2019 年因 AMD 致中度视力障碍患者占 4.56%，重度视力障碍者占 4.14%[2]。早期 AMD 与玻璃膜疣沉积有关，晚期分干性、湿性 2 种，其发病机制复杂，治疗方法有限，诊疗效果不佳[3-4]。越来越多的证据表明，AMD 与衰老和能量代谢密切相关，视网膜色素上皮（retinal pigment epithelial，RPE）细胞利用自噬过程释放代谢能量，使成分降解为简单分子，并产生维持能量稳态的三磷酸腺苷（adenosine triphosphate，ATP），随着 RPE 细胞的衰老，自噬过程减少，代谢能量降低，代谢碎片和蛋白质聚集体积累，从而导致 AMD[5-6]。研究也证明改善能量代谢可以延缓衰老，因此其可能是治疗 AMD 的一种潜在方法[7]。

随着快速、高通量转录组学测序方法的普及，对现有的高通量测序数据进行挖掘可为医学研究提供有效信息。Bulk RNA 数据基于细胞群体水平获得组织的平均特性，反映细胞群体的平均差异，而 Single-cell RNA 通过单个细胞的 mRNA 进行转录组分析，识别不同细胞亚群的细胞类型、状态和谱系，绘制细胞图谱、鉴定细胞异质性，可以精准解析细胞亚类功能及对疾病的影响[8-9]。本研究结合 Bulk RNA 测序和 Single-cell RNA 测序数据探究 AMD 发病机制中能量代谢关键基因并对潜在的调控中药化合物进行预测，为 AMD 药物靶点和治疗药物的研究提供新视角[10]。

1 资料与方法

1.1 Bulk RNA 和 Single-cell RNA 测序数据获取、处理及差异基因筛选

借助 GEO 数据库（https://www.ncbi.nlm.nih.gov/gds/）以“age-related macular degeneration”“AMD”作为检索词，种属选择“human sapiens”，查找下载 Bulk RNAseq 数据集 GSE135092[11]。该数据集共 537 个样本，其中来自视网膜组织 271 例；来自 RPE 组织 266 例。选取数据集中 266 例 RPE 组织样本［AMD 组（AMD 患者）RPE 组织黄斑区 26 例，非黄斑区 23 例；对照组（正常 RPE 组织）黄斑区 105 例，非黄斑区 112 例］纳入本研究分析。基于 R 软件“DESeq2”包［R 软件命令行输入 install.packages（“DESeq2”）后运行］对数据进行标准化和分析，以校正后 $P<0.05$ 为阈值，且 $|\log_2 FC|>2$［FC 表示差异倍数（fold change）］为初筛条件，对表达数据进行差异表达分析。

同样查找并下载 AMD Single-cell RNA 数据集 GSE188280[12]，该数据集包含 2 例 AMD 患者和 1 例正常人（对照）的视网膜和 RPE 组织的黄斑区和非黄斑区样本，选取 6 例 RPE 组织样本纳入本研究。使用 R 软件“Seurat”包［R 软件命令行依次输入 if (!requireNamespace ("BiocManager", quietly＝TRUE))和 install.packages ("BiocManager") 和 Bioc

Manager::install ("Seurat") 后运行]将 GSE188280 的表达矩阵创建为 Seurat 对象。过滤线粒体基因含量＞20%的细胞以及 feature＜500 或＞6 000 的细胞，仅选取 RPE 组织的细胞，最终获得 32 714 个细胞的转录组数据。通过“NormalizeData”函数对 GSE188280 的测序深度进行标准化，调用“FindVariableFeatures”函数使用“vst”方法检测数据集的 2 000 个可变特征，并使用“ScaleData”对数据进行缩放以排除测序深度的影响。

1.2 Bulk RNA 测序中 AMD 相关的能量代谢相关基因（energy metabolism-related genes，EMRGs）的蛋白相互作用（protein-protein interaction networks，PPI）、转录因子（transfer factor，TF）、微小 RNA（microRNA，miRNA）网络构建

EMRGs 来自 Gene Set Enrichment Analysis 数据库（GSEA，https://www.gsea-msigdb.org/），最终有 48 个 AMD 相关 EMRGs 从 GSE135092 的转录组数据集中提取出来。使用 Wilcoxcon 检验比较 GSE135092 RPE 组织黄斑区和非黄斑区来源的 EMRGs 的表达，P＜0.05 被认为具有统计学差异，使用 R 软件“pheatmap”包绘制热图，并用 Pearson 分析不同 EMRGs 之间的相关性。

借助 STRING（http://www.string-db.org/）在线工具和 Cytoscape 软件将差异表达的 EMRGs 上传构建 PPI 网络。进一步分析其 TF 和 miRNA 作用网络，将差异表达 EMRGs 上传至 Network Analyst（https://www.networkanalyst.ca/），参考 TF databases ENCODE（http://cistrome.org/BETA/）和 miRTarBase v8.0（https://mirtarbase.cuhk.edu.cn）数据库，分别使用 TF-gene Interactions 和 Gene-miRNA Interactions 模块进行分析。

1.3 Logistic 回归分析评估 AMD 患者 EMRGs 诊断能力

Logistic 回归模型正是作为被解释变量与其他自变量相互关系的一种线性模型，广泛用于医学领域[13]。受试者工作特征曲线（receiver operating charateristic，ROC）作为疾病诊断的常用统计学分析方法，是一种全面、准确评价诊断指标的工具[14]。将有差异的 EMRGs 作为自变量，将 AMD 发生情况作为因变量，使用 Logistic 回归分析差异表达 EMRGs 与是否患有 AMD 的关联性，并根据 Logistic 模型拟合结果利用 R 软件的“pROC”包绘制 ROC 曲线，并计算曲线下面积（area under curve，AUC），评估 AMD 患者 EMRGs 诊断能力。AUC＝0.5 时，说明关键靶基因无诊断价值；AUC＞0.5 时，面积越接近于 1，说明关键靶基因诊断的准确性越高；0.5＜AUC≤0.7，说明关键靶基因准确性较低；0.7＜AUC≤0.9 时，说明关键靶基因有准确性中等；AUC＞0.9 时，说明关键靶基因准确性高[15]。

1.4 最小绝对值收敛和选择算子（least absolute shrinkage and selection operator，LASSO）筛选 AMD 诊断标志物及风险评分构建

LASSO 算法是一种既能对变量进行选择和压缩来进行特征选择，又能得出参数估计值的基于线性回归模型的方法，可显著降低特征空间维度，有效地避免过拟合和多重共线性问题，被广泛应用和推广[16]。使用 LASSO 回归在差异表达基因中筛选诊断标志物。使用“glmnet”包的“glmnet”函数进行变量筛选，采用“cv.glmnet”函数进行交叉验证，然后筛选出使得变异系数最小的预后标志物组合。每个样本的风险评分的计算公式如下。用 Wilcoxcon 检验比较 AMD 和对照的风险评分，以箱式图展示。

$$\text{风险评分}=\sum_{i=1}^{n}\text{Coefficient}_i \times \text{mRNA Expression (gene}_i\text{)}$$

Coefficient 为 LASSO 回归系数，mRNA Expressiopn 为该基因的表达水平（$\log_2$ 转换）

1.5 Single-cell RNA 测序分析及细胞类型注释

主成分分析法（principal component analysis，PCA）是一种将高维空间复杂数据转化为低维空间的常用线性降维方法[17]。均匀流形逼近和投影（uniform manifold approximation and projection，UMAP）是基于流形学习技术和拓扑分析思想的降维算法，耗时短、对嵌入维数没有限制，具有可视化的优点，可用于医学、生物学数据分析[18]。研究应用 PCA 识别显著主成分（principal component，PC），用“Elbowplot”函数可视化 P 值分布。然后选 22 个 PC 进行 UMAP 分析以进行降维，用“FindNeighbors”默认参数以及 22 个 PC 维度参数来构建 PCA 空间中欧几里得距离的 K-最近邻域。然后通过调用“FindClusters”和“clustree”函数寻找到最佳分辨率 0.5，将细胞分为 19 个不同的簇。最后用“RunUMAP”函数进行降维以便探索数据集和可视化。

通过文献中已报道在 RPE 和各种脉络膜细胞类型中表达的基因列表以及人工注释进行本研究的细胞类型注释，例如，T 细胞的标记基因有 C 型蛋白酪氨酸磷酸酶受体（protein tyrosine phosphatase

receptor type C，*PTPRC*）、*CD3D*（CD3 delta）、*CD2*（CD2 molecule），B 细胞的标记基因有跨膜 4 域亚家族 A1（membrane spanning 4-domains A1，*MS4A1*）、*CD79A*，NK 细胞的标记基因有自然杀伤细胞颗粒蛋白 7（natural killer cell granule protein 7，*NKG7*）、颗粒酶 A（granzyme A，*GZMA*），骨髓细胞的标记基因有溶菌酶（lysozyme，*LYZ*）、*CD68*、Fc 区 III 受体 γ（Fc gamma RIIIA，*FCGR3A*）、*CD14*，内皮细胞的标记基因有血小板内皮细胞粘附分子 1（platelet endothelial cell adhesion molecule 1，*PECAM1*）、血管性血友病因子（von willebrand factor，*VWF*），成纤维细胞的标记基因有 α1-1 型胶原（collagen, type I, alpha 1，*COL1A1*）、*COL1A2*、核心蛋白聚糖（decorin，*DCN*）、胰岛素样生长因子 2（insulin-like growth factor 2，*IGF2*）、纤维蛋白 1（fibulin 1，*FBLN1*），施万细胞的标记基因有髓鞘质蛋白 1（proteolipid protein 1，*PLP1*），黑色素细胞的标记基因有黑色素 A（melan-A，*MLANA*）、黑素细胞蛋白（premelanosome protein，*PMEL*）、酪氨酸酶相关蛋白 1（tyrosinase related protein 1，*TYRP1*），巨噬细胞的标记基因有同种异体移植炎症因子 1（allograft inflammatory factor 1，*AIF1*），肥大细胞的标记基因有原癌基因 KIT（KIT proto-oncogene，*KIT*），平滑肌细胞的标记基因有蛋白信号调节因子 5（regulators of G-protein signaling 5，*RGS5*）、α 肌动蛋白 2（actin alpha 2，*ACTA2*）。

1.6 细胞亚群中 EMRGs 的表达

通过“DoHeatmap”函数展现 EMRGs 在 GSE188280 数据集不同细胞类型的表达情况，用热图展示。然后使用“DotPlot”函数绘制 6 个诊断标志物在不同细胞类型间的表达水平。最后使用“VlnPlot”函数绘制诊断标志物在不同细胞类型的表达情况的小提琴图。

1.7 调控中药预测及计算机分子对接、分子动力学模拟

Coremine Medical 数据库（https://www.pubgene.com/coremine-medical/）是国际上最先进的医学信息检索平台，可查询基因所影响的生物学过程[19]。通过 Coremine Medical 数据库，以 $P<0.05$ 为阈值，下载对所筛选的关键靶基因具有统计学意义的相关中药。中药系统药理学与分析平台 TCMSP（Traditional Chinese Medicine Systems Pharmacology Database and Analysis Platform）数据库（https://www.tcmsp-e.com/）能筛选中药化学成分，查找 499 味中药口服生物利用度、药物相似度等药动学特性[20]，通过 TCMSP 查找关键中药主要成分，从化学分子数据库 PubChem（https://pubchem.ncbi.nlm.nih.gov/）获取中药活性小分子成分配体的 3D 结构。从蛋白质数据库 UniProt（https://www.uniprot.org/）查询基因蛋白 Entry 号，RCSB PDB（Protein Data Bank）数据库（https://www.rcsb.org/）中下载关键基因对应蛋白受体 3D 结构。将中药小分子配体及关键基因对应蛋白受体运用 AutoDock Tools-1.5.6 软件进行分子对接及分子动力学模拟。

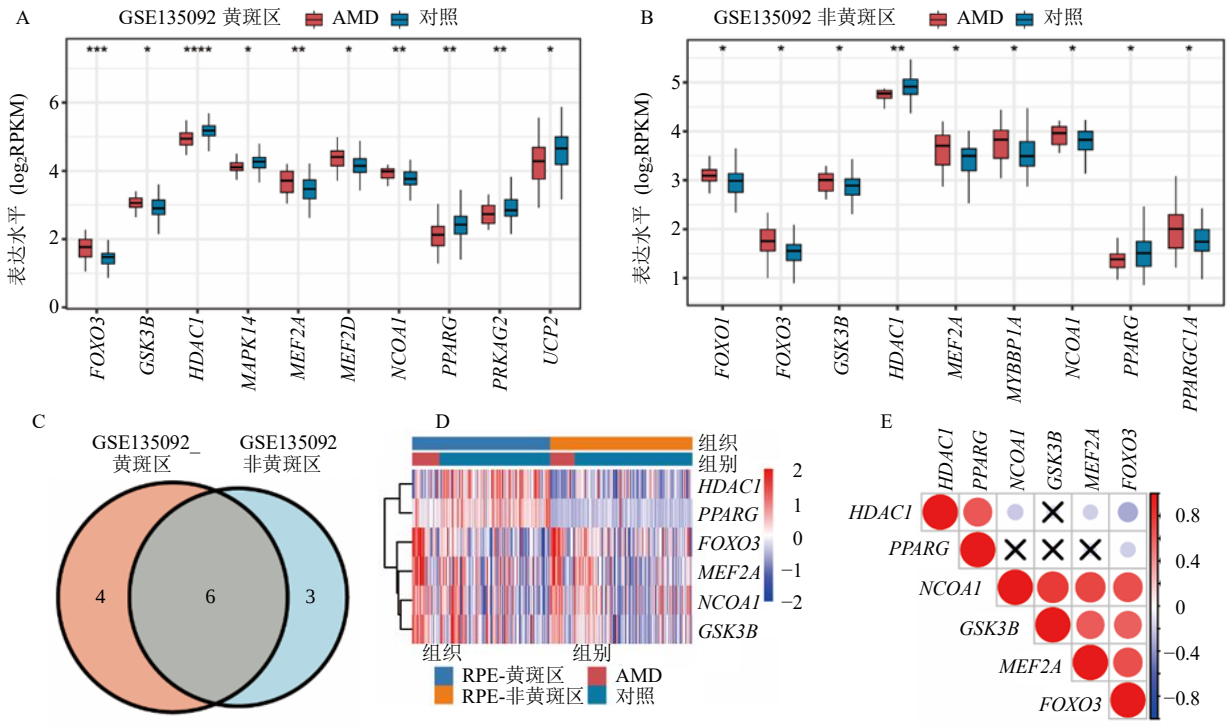
2 结果

2.1 Bulk RNA 测序筛选差异表达 AMD 相关 EMRGs 基因

研究比较 GSE135092 数据集中 AMD 组和对照组的黄斑区和非黄斑区组织 EMRGs 的表达水平。组间差异采用 Wilcoxcon 检验，如图 1-A 所示，在黄斑区发现了 10 个 EMRGs 的表达都具有显著差异（$P<0.05$）；如图 1-B 所示，在非黄斑区发现了 9 个 EMRGs 的表达具有显著差异（$P<0.05$）。如图 1-C 所示，两类组织差异表达基因中有 6 个交集基因，其中叉头蛋白 O3（forkhead box O3，*FOXO3*）、丝氨酸苏氨酸激酶 3β（glycogen synthase kinase 3 beta，*GSK3B*）、肌细胞增强因子 2A（myocyte enhancer factor 2A，*MEF2A*）、核受体共激活因子 1（nuclear receptor coactivator 1，*NCOA1*）在 AMD 组高表达，组蛋白脱乙酰酶 1（histone deacetylase 1，*HDAC1*）、过氧化物酶体增生激活受体 γ（peroxisome proliferator-activated receptor gamma，*PPARG*）在 AMD 组低表达。如图 1-D 所示，基因表达水平颜色从浅到深表示表达水平升高，蓝色表示呈负相关，红色表示呈正相关。EMRGs 表达水平的相关性矩阵如图 1-E 所示，*FOXO3* 与 *NCOA1*、*GSK3B*、*MEF2A* 呈显著正相关，*MEF2A* 与 *NCOA1*、*GSK3B* 呈显著正相关，*GSK3B* 和 *NCOA1* 以及 *PPARG* 和 *HDAC1* 分别呈显著的正相关；而 *FOXO3* 和 *HDAC1* 呈显著的负相关；红色表示正相关，蓝色表示负相关，颜色越深表示相关程度越高。对差异表达 EMRGs 之间的相关性进行分析，相关系数大于 0.65 的相关性结果的散点图如图 2-A～E 所示。

2.2 差异表达 AMD 相关 EMRGs 的 PPI、TF、miRNA 网络

将 6 个差异表达 AMD 相关的 EMRGs 进行 PPI



A-AMD 组和对照组在黄斑区的 EMRGs 表达水平箱式图；B-AMD 组和对照组在非黄斑区的 EMRGs 的表达水平箱式图；C-黄斑区和非黄斑区显著差异 EMRGs 的韦恩图；D-差异表达 EMRGs 在不同样本间的表达热图；E-差异表达 EMRGs 之间的相关性矩阵；黑色“X”表示不具有统计学相关性。

A-box plot of EMRGs expression level in the macular area of AMD and control groups; B-box plot of EMRGs expression level in the non-macular area of AMD and control groups; C-Venn map of EMRGs with significant differences between the macular and non-macular area; D-expression heat map of differentially expressed EMRGs between different samples; E-correlation matrix between differentially expressed EMRGs; the black “X” indicates no statistical correlation.

图 1　GSE135092 数据集 EMRGs 的表达情况及相关性分析

Fig. 1　Expression and correlation analysis of EMRGs in GSE135092 dataset

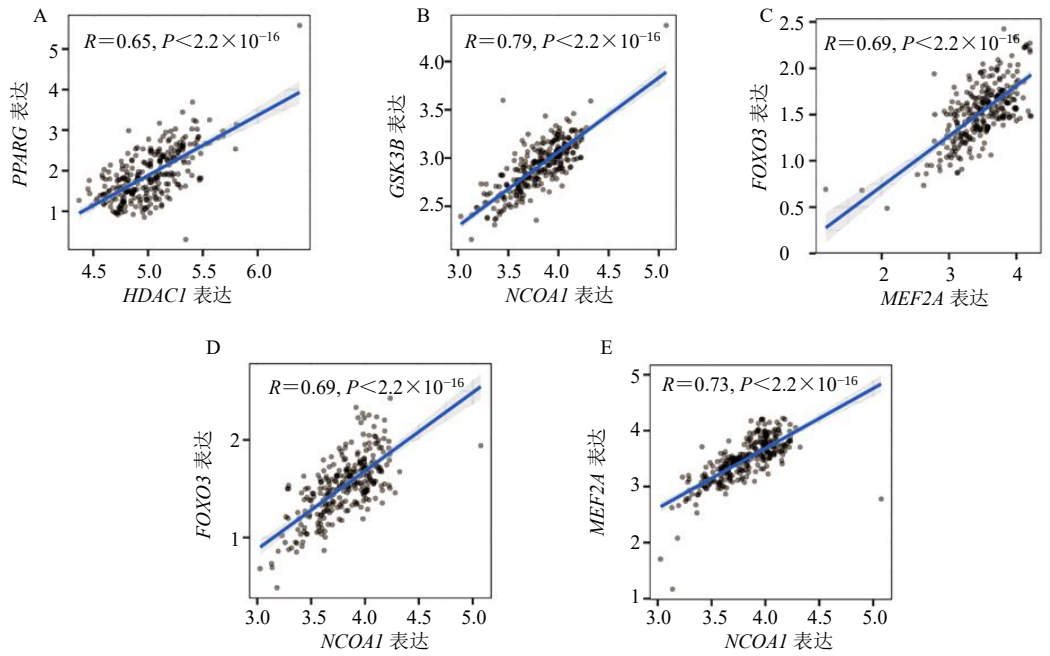
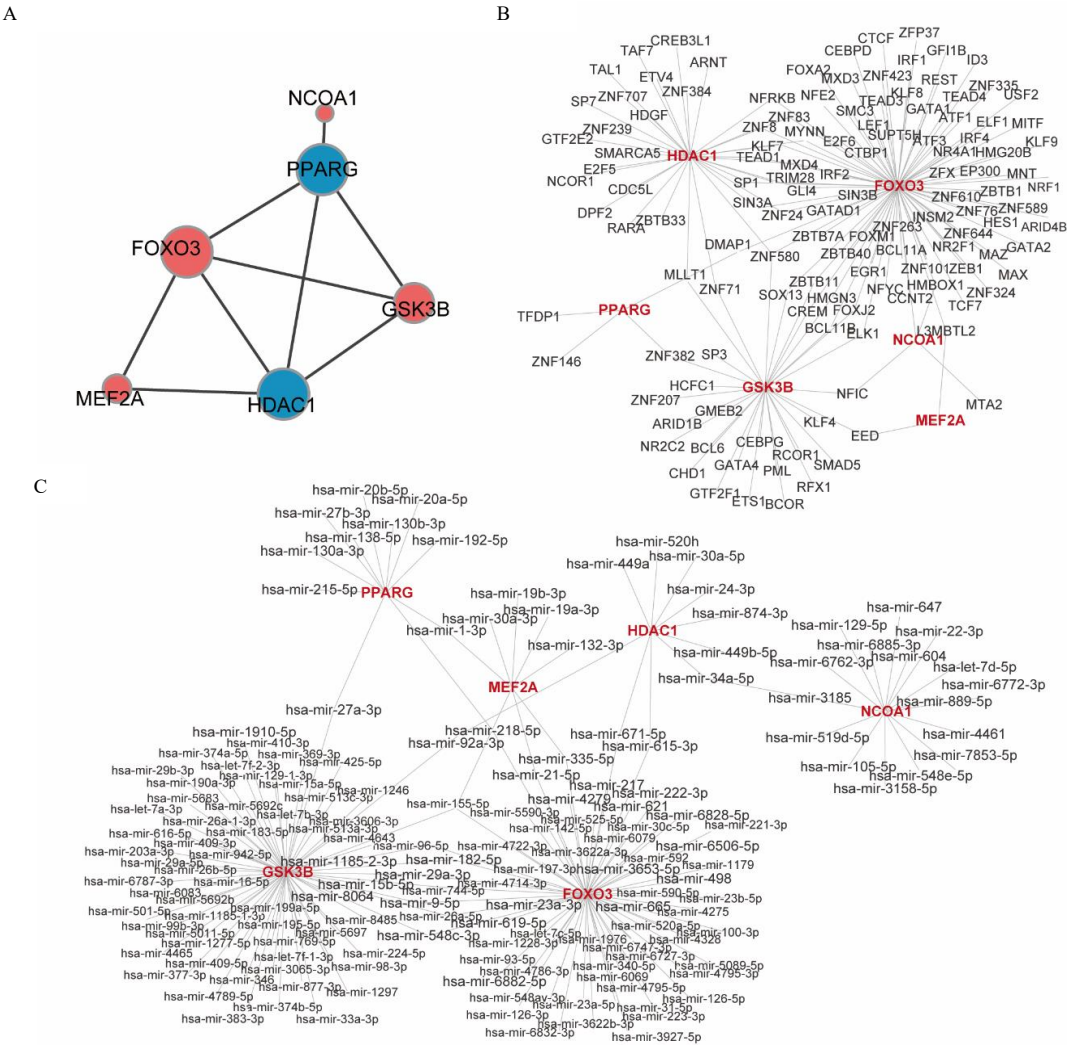


图 2　差异表达 EMRGs 相关性分析

Fig. 2　Correlation analysis of differentially expressed EMRGs

网络分析，PPI 网络如图 3-A 所示，红色表示上调，蓝色表示下调，节点越大表示和其他基因联系越多。mRNA-TF 互作结果显示，*FOXO3* 与 85 种 TF、*GSK3B* 与 31 种 TF、*HDAC1* 与 33 种 TF、*MEF2A* 与 2 种 TF、*NCOA1* 与 3 种 TF、*PPARG* 与 4 种 TF 存在互作关系，如图 3-B 所示。mRNA-miRNA 分析发现，*FOXO3* 与 64 种 miRNA、*GSK3B* 与 68 种 miRNA、*HDAC1* 与 10 种 miRNA、*MEF2A* 与 8 种 miRNA、*NCOA1* 与 17 种 miRNA、*PPARG* 与 11 种 miRNA 存在互作关系，如图 3-C 所示。



A-PPI 网络；B-mRNA 和 TF 互作网络；C-mRNA 和 miRNA 互作网络。

A-PPI network; B-mRNA and TF interaction network; C-mRNA and miRNA interaction network.

图 3　差异表达 EMRGs 互作网络分析

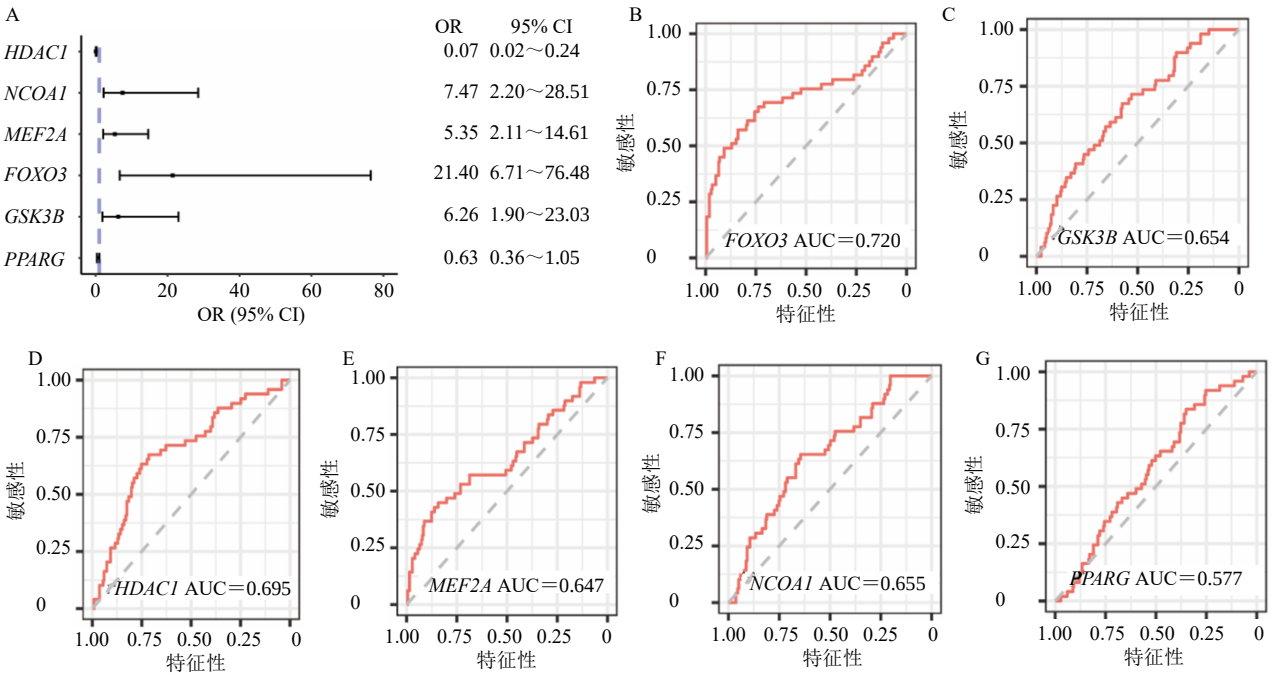
Fig. 3　Interaction network analysis of differentially expressed EMRGs

2.3　差异表达 AMD 相关 EMRGs 诊断 AMD 能力评估

使用 Logistic 回归分析差异表达 AMD 相关 EMRGs 与是否患有 AMD 的关联性，发现 *NCOA1*、*MEF2A*、*FOXO3*、*GSK3B* 表达分别与 AMD 发生呈正相关，*HDAC1* 表达与 AMD 发生呈负相关，如图 4-A 所示。根据 Logistic 模型拟合结果分别绘制了 ROC 曲线，AUC 分别为 *FOXO3* 0.720、*GSK3B* 0.654、*HDAC1* 0.695、*MEF2A* 0.647、*NCOA1* 0.655、*PPARG* 0.577，如图 4-B～G。

2.4　LASSO 分析差异表达 EMRGs 筛选 AMD 诊断标志物及风险评分

对获得的差异表达 EMRGs 进行 LASSO 回归分析，并利用具有 10 倍交叉验证的部分似然偏差来计算最佳 Lambda 值（λ），其中红点代表每个 λ 的目标参量，虚线与横坐标相交处为最优 λ 值。如图 5-A 所示，当均方误差取最小值时，预测模型描述

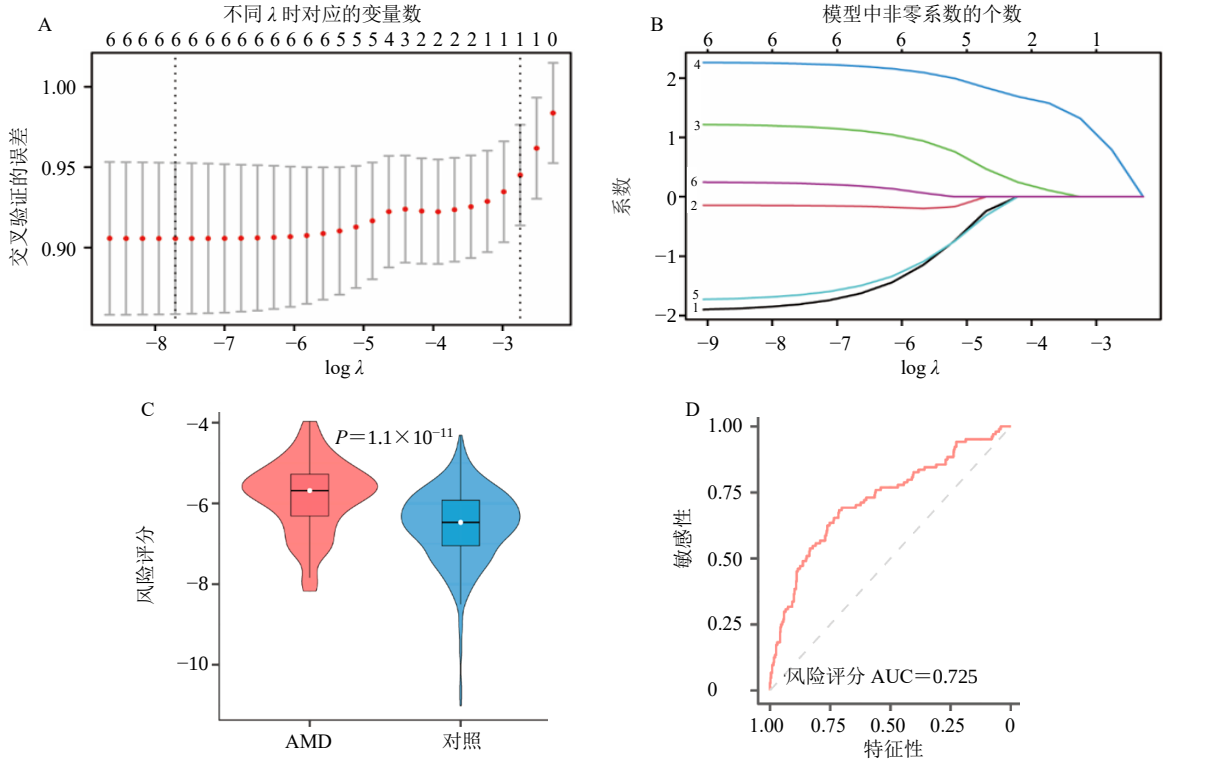


A-差异表达 EMRGs Logistic 回归分析森林图；B-*FOXO3* 的 ROC 曲线；C-*GSK3B* 的 ROC 曲线；D-*HDAC1* 的 ROC 曲线；E-*MEF2A* 的 ROC 曲线；F-*NCOA1* 的 ROC 曲线；G-*PPARG* 的 ROC 曲线。

A-Logistic regression analysis forest map of differentially expressed EMRGs; B-ROC curve of *FOXO3*; C-ROC curve of *GSK3B*; D-ROC curve of *HDAC1*; E-ROC curve of *MEF2A*; F-ROC curve of *NCOA1*; G-ROC curve of *PPARG*.

图 4　差异表达 EMRGs Logistic 回归分析与 ROC 曲线

Fig. 4　Logistic regression analysis and ROC curve of differentially expressed EMRGs



A-LASSO 回归的特征变量选择；B-LASSO 回归拟合系数轨迹；C-AMD 和对照组的风险评分比较；D-风险评分构建诊断模型的 ROC 曲线。

A-feature variable selection of LASSO regression; B-fitting coefficient trajectory of LASSO regression; C-comparison of risk scores between AMD and control groups; D-ROC curve for constructing diagnostic models based on risk scores.

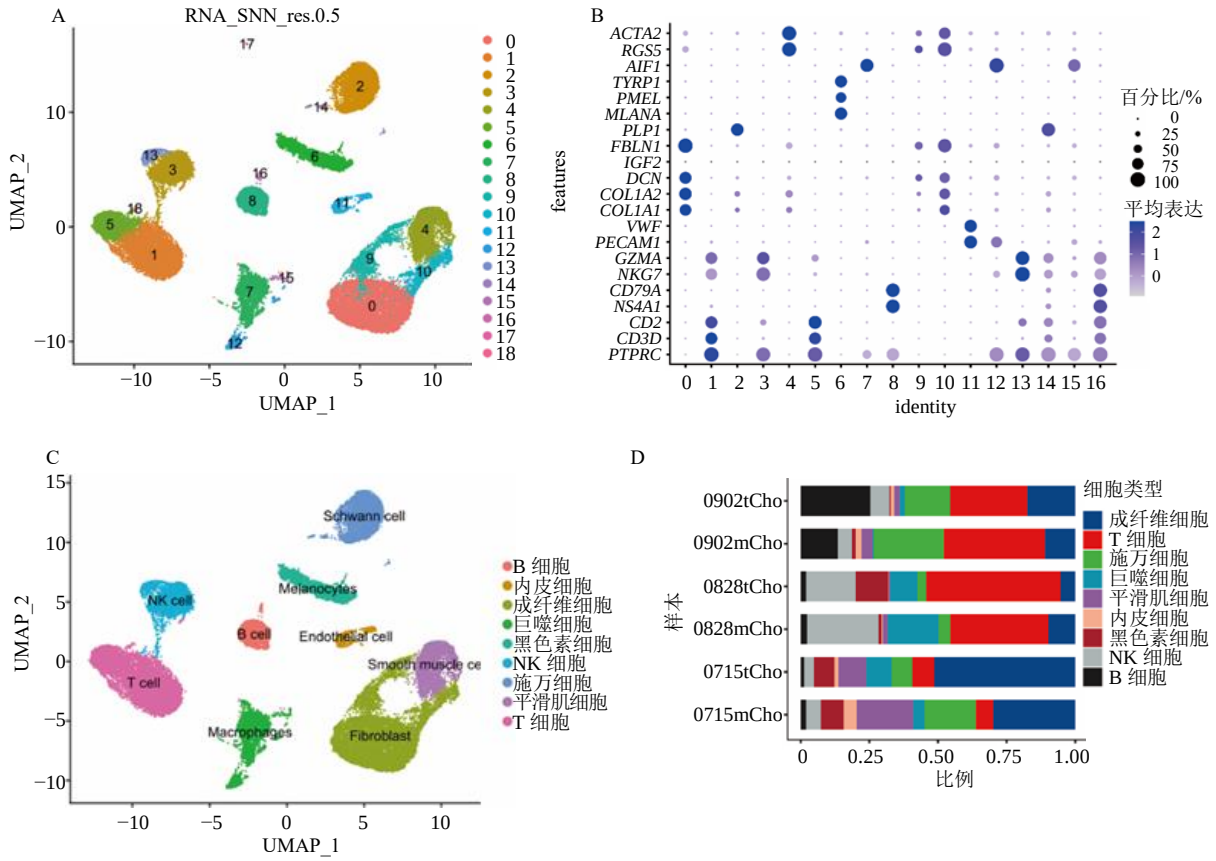
图 5　LASSO 回归筛选诊断 AMD 相关 EMRGs

Fig. 5　LASSO regression screening for diagnostic AMD-related EMRGs

实验数据精确度越高，对应的 λ 值即可确定 LASSO 回归复杂度调整的程度。如图 5-B 所示，6 个差异 EMRGs 都被筛选出来作为诊断标志物。根据 LASSO 回归模型的分析结果确定候选预后标志物的系数，计算风险评分，风险评分＝−1.821 *HDAC1*－0.144 *NCOA1*＋1.185 *MEF2A*＋2.243 *FOXO3*－1.662 *GSK3B*＋0.229 *PPARG*。分别计算 AMD 和对照组样本的风险评分，组间差异采用 Wilcoxcon 检验，结果表明 AMD 组的风险评分显著高于对照组，如图 5-C 所示。采用 Logistic 回归模型构建诊断模型，模型拟合结果显示 ROC AUC 为 0.725，如图 5-D 所示。

2.5　Single-cell RNA 测序分析 RPE 组织的细胞异质性

本研究对单细胞 GSE188280 数据集进行质控，过滤线粒体基因含量＞20%，feature＜500 或＞6 000 的细胞后获得了 105 424 个细胞，提取来源 RPE 组织的细胞共计 32 714 个，以供下游分析。应用 UMAP 降维进行可视化，将 32 714 个细胞成功分类为 19 个独立的簇，如图 6-A、B 所示，颜色越深表示表达水平越高，圆圈越大表示基因在细胞群内的表达比例越高。使用 marker 基因将细胞簇鉴定为共 9 种细胞类型，如图 6-C 所示，cluster 1、5 注释为 T 细胞（7 481，22.92%）；cluster 8、16 注释为 B 细胞（1 884，5.77%）；cluster 11 注释为内皮细胞（593，1.81%）；cluster 0、9、10 注释为成纤维细胞（8 773，26.88%）；cluster 7、12、15 注释为巨噬细胞（2 433，7.45%）；cluster 6 注释为黑色素细胞（2 049，6.27%）；cluster 3、13 注释为 NK 细胞（3 016，9.24%）；cluster 2、14 注释为施万细胞（3 766，11.54%）；cluster 4 注释为平滑肌细胞（2 634，8.07%）。每个样本间的各细胞比例如图 6-D 所示。



A-32 714 个 RPE 组织细胞通过 UMAP 聚类为 19 个细胞簇；B-marker 基因在各细胞簇的表达情况气泡图；C-通过 marker 基因将细胞注释为 9 种细胞类型；D-百分比堆叠图展现每个样本的 9 种细胞比例；0902、0715 为 AMD，0828 为对照，mCho 表示 RPE 组织黄斑区，tCho 表示 RPE 组织非黄斑区。

A-32 714 RPE tissue cells were clustered into 19 cell clusters through UMAP; B-bubble diagram of the expression of marker gene in various cell clusters; C-cells are annotated into nine cell subsets through marker gene; D-percentage stacking chart displays the nine cell ratios of each sample; 0902 and 0715 are AMD, 0828 is control, mCho represents macular area of RPE tissue, tCho represents non-macular area of RPE tissue.

图 6　单细胞 RNA 测序数据集 GSE188280 鉴定 RPE 组织细胞亚群

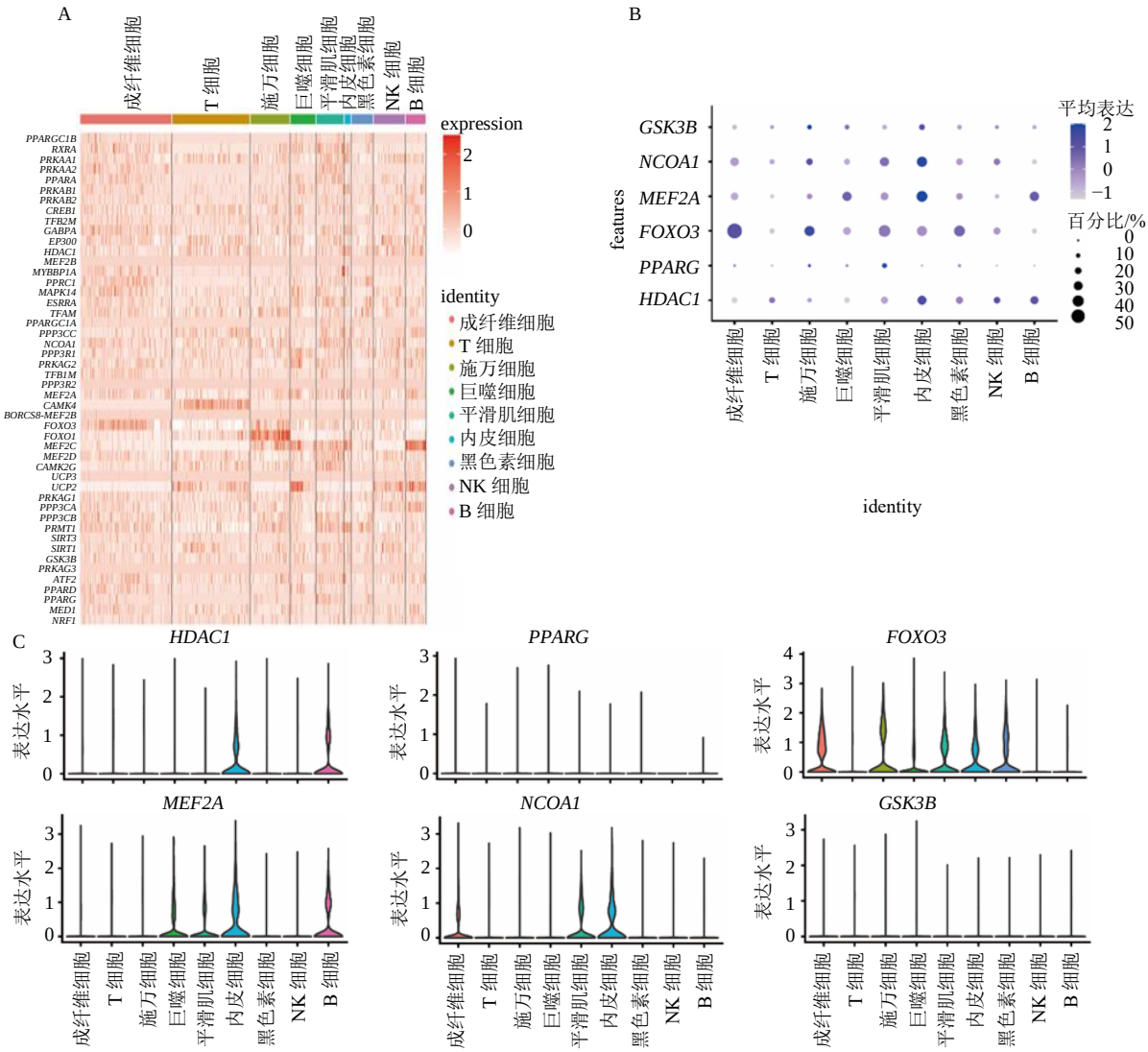
Fig. 6　Identification of RPE cell subsets using single-cell RNA sequencing dataset GSE188280

2.6 RPE 组织细胞亚群中 EMRGs 的表达

利用热图展现 48 个 EMRGs 在各类型细胞中的表达情况，如图 7-A 所示。其中 6 个差异 EMRGs 在各类型细胞中的表达如图 7-B 所示，颜色越深表示表达水平越高，圆圈越大表示基因在细胞群内的表达比例越高；6 个差异 EMRGs 在各类型细胞中的表达小提琴图如图 7-C 所示。*HDAC1*、*MEF2A* 在内皮细胞和 B 细胞中存在较高表达，*FOXO3* 在内皮细胞、成纤维细胞、施万细胞、平滑肌细胞、巨噬细胞中都存在较高表达，*NCOA1* 在内皮细胞中存在较高表达。

2.7 潜在中药化合物筛选与计算机分子对接、分子动力学模拟

从 Coremine Medical 数据库，以 $P<0.05$ 为阈值，筛选出与关键靶基因具有统计学意义的相关中药黄芪、人参、山药、山茱萸、桑白皮、桑叶、淫羊藿、西红花和丹参。从 TCMSP 数据库根据中药化学物质口服生物利用度（oral bioavailability，OB）、类药性（drug-likeness，DL）筛选关键中药主要活性物质 285 种，选择生物活性最高的 9 个活性物质（丹参醇 A、丹参醇 B、薯蓣皂素、富马碱、马卡因、白桦脂酸、马尔肯久纳醇苯酯、正庚醛、淫羊藿 A）及对应的关键基因靶点蛋白（FOXO3、HDAC1、MEF2A）进行分子对接。所有的对接实验运用 AutoDock Tools-1.5.6 软件，并根据结合自由能评估最终对接结构，最佳前 4 组利用 Pymol 2.1 软件进行可视化，如图 8-A～D 所示。其中，丹参醇 B 能够与蛋白 HDAC1 的 GLY-149

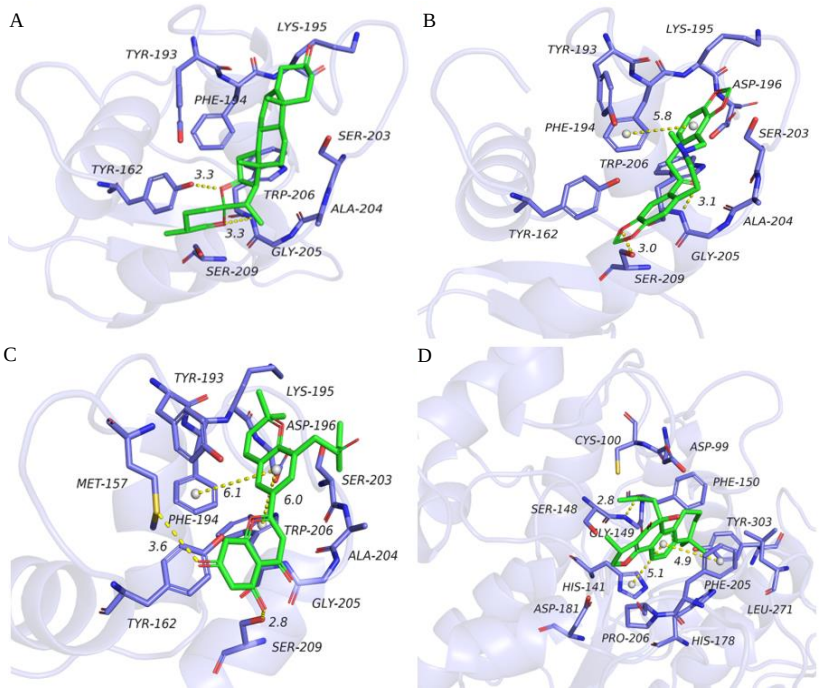


A-48 个 EMRGs 在细胞亚群表达的热图；B-6 个差异表达 EMRGs 在细胞亚群表达的气泡图；C-6 个差异表达 EMRGs 在细胞亚群表达的小提琴图。

A-heat map of 48 EMRGs expressed in cell subsets; B-bubble plot of six differentially expressed EMRGs expressed in cell subsets; C-violin diagram of six differentially expressed EMRGs expressed in cell subsets.

图 7 差异表达 EMRGs 在单细胞 RNA 测序数据集 GSE188280 中的表达水平

Fig. 7 Expression level of differentially expressed EMRGs in single-cell RNA sequencing dataset GSE188280



A-FOXO3 和薯蓣皂苷元分子对接图；B-FOXO3 和延胡索碱分子对接图；C-FOXO3 和淫羊藿 A 分子对接图；D-HDAC1 和丹参醇 B 分子对接图。A-molecular docking of FOXO3 and diosgenin; B-molecular docking of FOXO3 and fumarine; C-molecular docking of FOXO3 and yinyanghuo A; D-molecular docking of HDAC1 and danshenol B.

图 8　分子对接可视化结果

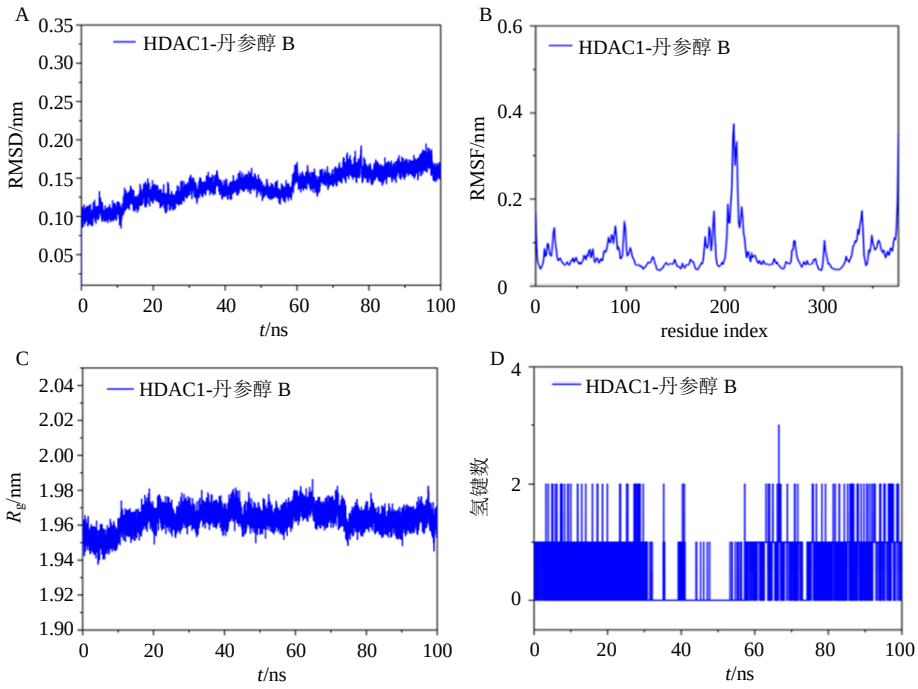
Fig. 8　Visualization results of molecular docking

氨基酸形成氢键距离短（0.28 nm），结合能力强；还能与蛋白口袋的 PHE-205、HIS-141、PHE-150、PRO-206、LEU-271、TYR-303 形成很好的疏水作用，特别是化合物的苯环能够与 PHE-205、HIS-141 形成 p-p 共轭作用，对稳定小分子有着重要贡献，如图 8-D 所示。

借助 Gromacs 2020 软件包对分子对接最佳的丹参醇 B 与 HDAC1 蛋白进行分子动力学模拟，使用 Gromacs 2020 程序的 g_MMPBSA 方法对结合自由能进行计算。结果显示，复合物的平均方根偏差（RMSD）均小于 0.2 nm，并且基本在 60 ns 左右达到动态平衡，表明小分子与靶点蛋白匹配较好能形成稳定的复合物，如图 9-A 所示。由平均方根波动（RMSF）结果可知，蛋白与小分子作用形成的复合物中少部分氨基酸构象变化较大，大部分变化较小，说明小分子与蛋白相互作用形成稳定的复合物，如图 9-B 所示。HDAC1 蛋白的回转半径（R_g）在前 20 ns 略有升高，之后缓慢下降，蛋白与化合物的结合促使蛋白维持了更多的疏水接触，内部形成更多有效相互作用，而此时蛋白空腔扩大以更好地匹配化合物，从而促进复合物的稳定性，如图 9-C 所示。统计蛋白与小分子的氢键数量发现，化合物均能够与蛋白口袋氨基酸形成一个氢键相互作用，这些氢键在稳定的小分子与蛋白结合方面起到了重要作用，如图 9-D 所示。丹参醇 B 与 HDAC1 蛋白的结合自由能为（−75.780±11.918）kJ/mol，范德华力发挥主要作用[（−149.094±9.569）kJ/mol]。氢键使静电相互作用在稳定小分子方面也存在贡献[（−23.388±6.477）kJ/mol]。综上，丹参醇 B 能够与 HDAC1 蛋白有着很强的亲和力，可以促使小分子与蛋白形成为稳定的复合物，从而发挥活性作用。

3　讨论

RPE 是视网膜的基本组成部分，RPE 细胞具有高度的新陈代谢活性，其依赖能量代谢来完成运输营养、吞噬光感受器脱落的外节盘、消除活性氧、分泌因子等功能，对维持视网膜的健康至关重要，其功能异常将引起 AMD 等视网膜退行性疾病[21-22]。1 项研究比较从流产胎儿或老年供体眼球中分离出来的 RPE 细胞的能量代谢，发现线粒体耗氧量（ATP 生产）在老化细胞中完全丧失储备能力，而糖酵解在老化细胞中也被削弱，说明 RPE 细胞的能量代谢随年龄增加而减弱[23]。烟酰胺腺嘌



A-HDAC1 和丹参醇 B RMSD 图；B-HDAC1 和丹参醇 B RMSF 图；C-HDAC1 和丹参醇 B 回转半径图；D-HDAC1 和丹参醇 B 氢键数量图。
A-RMSD graph of HDAC1 and danshenol B; B-RMSF graph of HDAC1 and danshenol B; C-R_g graph of HDAC1 and danshenol B; D-H bond number diagram of HDAC1 and danshenol B.

图 9　分子动力学模拟可视化结果

Fig. 9　Visualization results of molecular dynamics simulation

吟二核苷酸（nicotinamide adenine dinucleotide，NAD^+）是一种中心代谢辅助因子，在调节细胞代谢和能量平衡方面起关键作用，有研究评估小鼠 RPE 中关键的 NAD^+生物合成基因的时间表达和相关的 NAD^+水平，发现 NAD^+水平随着年龄的增长而下降，也说明能量对 RPE 健康的重要性[24]。与健康对照组的 RPE 细胞相比，从 AMD 患者中分离出来的 RPE 细胞的能量代谢有障碍，该研究指出衰老加上外部压力扰乱 RPE 细胞的代谢是 AMD 重要病理特征[25]。总之，这些研究支持增强能量代谢对延缓衰老和保持 RPE 活力有潜在影响，或许是治疗 AMD 的有效方式。

高通量技术有低成本、测序快、通量高等优势，是新一代测序技术，其中转录组测序对疾病分子机制以及药物靶点的预测有一定帮助。本研究借助转录组测序数据，比较了 AMD 患者和非 AMD 患者的 mRNA 的表达谱，并使用相关计算方法筛选，并进行分子对接发现 *FOXO3* 和 *HDAC1* 是与 AMD 有较大关联的关键基因。研究发现 FOXOs 转录因子在细胞能量代谢、线粒体自噬等方面发挥着重要的调控作用，FOXO3 可调节葡萄糖代谢中的炎症标志物，FOXO3a 影响糖酵解、三羧酸循环中关键酶的表达[26-27]。FOXOs 主要通过能量稳态的中枢调节因子（adenosine monophosphate，AMP）依赖的蛋白激酶（adenosine 5′-monophosphate-activated protein kinase，AMPK）参与能量代谢，当能量缺失后细胞内的高 AMP/ATP 比率会导致 AMPK 激活，而活化的 AMPK 可以磷酸化 FOXO3 的 6 个不同氨基酸残基，导致抗氧化反应和能量利用通路的相关基因表达上调[28]。查阅 FOXO3 相关文献，未见与 AMD 相关报道，因此 FOXO3 或许是作用于 AMD 的新靶点。

HDAC 家族成员参与血管生成、炎症等生物过程，在 AMD 的病理生理学中发挥着重要作用[29]。在内皮细胞中，HDAC1 通过调节主要的内皮细胞功能，如血管生成、炎症信号、氧化还原平衡和一氧化氮信号，介导外部和环境刺激的影响。血管生成最常被内皮细胞 HDAC1 抑制，HDAC1 通过组蛋白和非组蛋白的去乙酰化对内皮功能的依赖性调节发挥作用[30]。HDAC1 表达减少血管新生增多，脉络膜新生血管是湿性 AMD 主要病理特征。相比于正常人，HDAC1、2、5 和 6 在 AMD 患者的视网膜中的表达减少[31]。结合 mRNA-TF、mRNA-miRNA

调控网络可进一步去探讨 AMD 发病机制。

本研究针对关键基因筛选治疗 AMD 的潜在中药化合物，得到黄芪、人参、山药、山茱萸、桑白皮、桑叶、淫羊藿、西红花和丹参共 9 味中药，主要归脾、肝、肾经，有补气健脾、清润明目、补阳活血之功。中药调控能量代谢主要通过葡萄糖代谢信号通路、线粒体能量代谢等途径来缓解病变，并且发现多数中药与脾、肝、肾关系密切[32]。脾为气血化生之源，气血是人体活动的物质基础，黄芪、人参、山药益气健脾，有助于水谷精微物质被输布散播至全身及目窍，发挥濡养作用。现代代谢组学研究指出，黄芪、人参可能通过转化生长因子 β1（transforming growth factor-β1，TGF-β1）/Notch 信号，改善线粒体能量信号，对线粒体动力学相关蛋白发挥作用[33]。肝开窍于目，肝气调达气血和顺，则目窍得养，肝气不舒，气郁化火，肝火上炎，耗伤精血，则目窍失养。桑叶清肝明目，研究发现桑叶参与单磷酸腺苷激活的 AMPK 能量代谢信号通路、胰岛素调控相关的信号通路等[34]。肾为命门之火，肾之阳又称元阳、真阳，温养五脏六腑官窍，肾阳虚衰，则脏腑官窍失于温养。淫羊藿温阳补肾，其发挥温补效应机制与淫羊藿苷调控 PTEN 诱导激酶 1（PTEN induced kinase 1，PINK1）/parkin 信号通路降低细胞线粒体自噬水平、改善能量代谢有关[35]。丹参活血通经，可以提高线粒体膜的完整性，降低其通透性，抑制大量钙离子涌入线粒体，保护参与代谢的酶类，在缺血缺氧时，促进三羧酸循环，维护呼吸链功能，增加线粒体 ATP 酶活性[36]。以上研究均支持所预测的健脾益气、清肝明目、温阳补肾、活血通经类中药可改善机体能量代谢，从而防治疾病。结合分子动力学模拟证明丹参有效成分丹参酮 B 与其对应基因转录蛋白 HDAC1 有着很强的亲和力，可促使小分子与蛋白形成为稳定的复合物而发挥活性作用，故丹参通过能量代谢过程对改善 AMD 具有潜在作用。因此，对这些中药化合物的关注可能会给 AMD 的防治带来新的研究方向。

综上所述，本研究通过 Bulk RNA 和 Single-cell RNA 测序数据探究 AMD 的能量代谢相关关键基因，并通过关键基因分析得到潜在治疗中药，借助分子对接技术及分子动力学模拟技术选出丹参酮 B 是最有潜在治疗价值的中药化合物，为 AMD 后续药物开发和临床应用提供一定的研究基础。但本研究仍然存在不足，单细胞测序较贵，尚未普及，AMD 相关基因芯片数量较少，除此之外本研究还需进一步体内或体外实验验证关键基因及相关中药的治疗效果。

利益冲突 所有作者均声明不存在利益冲突

参考文献

[1] Deng Y H, Qiao L F, Du M Y, *et al*. Age-related macular degeneration: Epidemiology, genetics, pathophysiology, diagnosis, and targeted therapy [J]. *Genes Dis*, 2021, 9(1): 62-79.

[2] 林艳辉, 高莉茉, 姜文敏. 基于 GBD 数据库中国年龄相关性黄斑变性流行病学负担分析 [J]. 中南大学学报: 医学版, 2023, 48(1): 106-113.

[3] Heesterbeek T J, Lechanteur Y T E, Lorés-Motta L, *et al*. Complement activation levels are related to disease stage in AMD [J]. *Invest Ophthalmol Vis Sci*, 2020, 61(3): 18.

[4] 邵毅, 温佳怡, 令倩. 年龄相关性黄斑变性诊断与治疗规范: 2022 年英国皇家眼科医学会指南解读 [J]. 眼科新进展, 2023, 43(2): 85-88.

[5] Yang J, Zhou R M, Ma Z Y. Autophagy and energy metabolism [J]. *Adv Exp Med Biol*, 2019, 1206: 329-357.

[6] Nashine S. Potential therapeutic candidates for age-related macular degeneration (AMD) [J]. *Cells*, 2021, 10(9): 2483.

[7] Feng Z Y, Hanson R W, Berger N A, *et al*. Reprogramming of energy metabolism as a driver of aging [J]. *Oncotarget*, 2016, 7(13): 15410-15420.

[8] Lähnemann D, Köster J, Szczurek E, *et al*. Eleven grand challenges in single-cell data science [J]. *Genome Biol*, 2020, 21(1): 31.

[9] Kulkarni A, Anderson A G, Merullo D P, *et al*. Beyond bulk: A review of single cell transcriptomics methodologies and applications [J]. *Curr Opin Biotechnol*, 2019, 58: 129-136.

[10] 安宁, 张福燕, 秦波. 单细胞转录组测序在年龄相关性黄斑变性研究中的应用 [J]. 国际眼科杂志, 2022, 22(6): 964-968.

[11] Orozco L D, Chen H H, Cox C, *et al*. Integration of eQTL and a single-cell atlas in the human eye identifies causal genes for age-related macular degeneration [J]. *Cell Rep*, 2020, 30(4): 1246-1259.e6.

[12] Zauhar R, Biber J, Jabri Y, *et al*. As in real estate, location matters: Cellular expression of complement varies between macular and peripheral regions of the retina and supporting tissues [J]. *Front Immunol*, 2022, 13: 895519.

[13] 蒋仕旗. Logistic 回归模型的一种改进弹性网估计 [J]. 数学理论与应用, 2022, 42(2): 108-119.

[14] Zweig M H, Campbell G. Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool

in clinical medicine [J]. *Clin Chem*, 1993, 39(4): 561-577.

[15] 陈卫中, 倪宗瓒, 潘晓平, 等. 用ROC曲线确定最佳临界点和可疑值范围 [J]. 现代预防医学, 2005, 32(7): 729-731.

[16] 孟金彪, 钱萌, 李存志, 等. 基于 Lasso 和模糊互信息多标记特征选择算法 [J]. 电子技术与软件工程, 2019(10): 186-188.

[17] 张碧颖, 苏海霞, 林倩玮, 等. 基于函数型主成分分析的稀疏纵向数据建模研究 [J]. 中国卫生统计, 2023, 40(2): 162-166.

[18] 易灿灿, 庹帅, 涂闪, 等. 基于UMAP辅助的模糊C聚类方法进行太赫兹光谱识别 [J]. 光谱学与光谱分析, 2022, 42(9): 2694-2701.

[19] Gorji-Bahri G, Moghimi H R, Hashemi A. RAB5A is associated with genes involved in exosome secretion: Integration of bioinformatics analysis and experimental validation [J]. *J Cell Biochem*, 2021, 122(3/4): 425-441.

[20] Ru J L, Li P, Wang J N, *et al*. TCMSP: A database of systems pharmacology for drug discovery from herbal medicines [J]. *J Cheminform*, 2014, 6: 13.

[21] Yam M, Engel A L, Wang Y K, *et al*. Proline mediates metabolic communication between retinal pigment epithelial cells and the retina [J]. *J Biol Chem*, 2019, 294(26): 10278-10289.

[22] Merle D A, Provenzano F, Jarboui M A, *et al*. mTOR inhibition via rapamycin treatment partially reverts the deficit in energy metabolism caused by FH loss in RPE cells [J]. *Antioxidants*, 2021, 10(12): 1944.

[23] Rohrer B, Bandyopadhyay M, Beeson C. Reduced metabolic capacity in aged primary retinal pigment epithelium (RPE) is correlated with increased susceptibility to oxidative stress [J]. *Adv Exp Med Biol*, 2016, 854: 793-798.

[24] Jadeja R N, Powell F L, Jones M A, *et al*. Loss of NAMPT in aging retinal pigment epithelium reduces NAD^+ availability and promotes cellular senescence [J]. *Aging*, 2018, 10(6): 1306-1323.

[25] Armento A, Honisch S, Panagiotakopoulou V, *et al*. Loss of complement factor H impairs antioxidant capacity and energy metabolism of human RPE cells [J]. *Sci Rep*, 2020, 10(1): 10320.

[26] 周欢, 孙冶, 刘春英, 等. 基于 FOXO3a 调控下的糖酵解途径探讨康艾注射液改善 A549/DDP 细胞顺铂耐药性研究 [J]. 辽宁中医药大学学报, 2022, 24(5): 13-18.

[27] Lundell L S, Massart J, Altıntaş A, *et al*. Regulation of glucose uptake and inflammation markers by FOXO1 and FOXO3 in skeletal muscle [J]. *Mol Metab*, 2019, 20: 79-88.

[28] 张仙宏, 魏萌萌, 袁冬冬, 等. 转录因子 FOXOs 家族调控肿瘤生物学功能的研究进展 [J]. 生理学报, 2022, 74(5): 843-855.

[29] Tang J H, Yan H D, Zhuang S G. Histone deacetylases as targets for treatment of multiple diseases [J]. *Clin Sci*, 2013, 124(11): 651-662.

[30] Dunaway L S, Pollock J S. HDAC1: An environmental sensor regulating endothelial function [J]. *Cardiovasc Res*, 2022, 118(8): 1885-1903.

[31] Noh S J, Jeong W J, Rho J H, *et al*. Sensitization of RPE cells by alpha B-crystallin siRNA to SAHA-induced stage 1 apoptosis through abolishing the association of alpha B-crystallin with HDAC1 in SC35 speckles [J]. *Invest Ophthalmol Vis Sci*, 2008, 49(11): 4753-4759.

[32] 郭园园, 郑琴, 胡鹏翼, 等. 中药调控能量代谢治疗神经退行性疾病的研究进展 [J]. 中药新药与临床药理, 2020, 31(11): 1384-1388.

[33] 黄小楼, 平兰芝, 谭晓波, 等. 益气活血药方对心肌缺血再灌注损伤模型大鼠氧化应激及线粒体能量代谢的影响 [J]. 河北中医, 2023, 45(4): 597-602.

[34] 黄青昕, 刘婷, 胡淇睿, 等. 基于体外模拟消化、发酵和网络药理学探讨桑叶酚类物质的潜在生物学活性 [J]. 中草药, 2023, 54(10): 3205-3215.

[35] 程相阁, 刘万周, 王东伟. 淫羊藿苷对心力衰竭大鼠线粒体自噬及能量代谢的影响 [J]. 西北药学杂志, 2022, 37(4): 51-57.

[36] 阎晓悦, 刘炜. 中药对线粒体能量代谢的影响 [J]. 中国实用医药, 2019, 14(9): 192-193.

[责任编辑 潘明佳]