

单细胞转录组学与机器学习综合分析 揭示腹主动脉瘤潜在生物标志物

王玉涛^{1,2}, 孙岩³

(1. 山东中医药大学第一临床医学院, 山东 济南 250355; 2. 济南市中医医院周围血管病科, 山东 济南 250012;

3. 山东第一医科大学附属省立医院血管外科, 山东 济南 250021)

摘要:目的 利用单细胞转录组测序(single-cell RNA sequencing, scRNA)数据分析、加权基因共表达网络分析(weighted gene co-expression network analysis, WGCNA)、机器学习算法和免疫浸润分析筛选腹主动脉瘤(abdominal aortic aneurysm, AAA)潜在的生物标志物。方法 下载基因表达数据库中包含AAA和正常主动脉(normal aorta control, NAC)的scRNA测序数据,经数据质量控制、降维、差异分析、细胞类型注释和拟时序分析后,筛选AAA发生过程中最早分化的细胞类型,筛选差异表达基因(differential expressed genes, DEGs),进行高维WGCNA(high dimensional WGCNA, hdWGCNA),识别与AAA相关的基因模块,并进行富集分析;下载包含AAA和NAC的常规转录组测序数据,进行差异分析、WGCNA,将scRNA样本DEGs、常规转录组DEGs和WGCNA结果进行整合,筛选与AAA病变相关的基因,并进行基因本体(gene ontology, GO)和京都基因与基因组百科全书(Kyoto Encyclopedia of Genes and Genomes, KEGG)信号通路富集分析。利用最小绝对收缩和选择算子(least absolute shrinkage and selection operator, LASSO)、支持向量机递归特征消除(support vector machine recursive feature elimination, SVM-RFE)和随机森林(random forest, RF)等机器学习方法筛选AAA的潜在生物标志物,并进行免疫浸润分析。结果 scRNA数据分析结果显示,内皮细胞是AAA发生过程中最早分化的细胞类型,共获得853个scRNA DEGs;hdWGCNA识别出与AAA相关的2个基因模块,显著富集于辅助性T细胞17细胞分化、辅助性T细胞1和2细胞分化等信号通路。常规转录组分析共获得162个DEGs;整合后获得17个AAA相关基因,显著富集于趋化因子、辅助性T细胞17细胞分化、辅助性T细胞1和2细胞分化等信号通路。机器学习算法识别出AAA的潜在生物标志物生态病毒整合位点2B(ecotropic viral integration site 2B, EVI2B)。EVI2B在AAA样本中的表达量高于NAC样本。免疫浸润结果显示,AAA样本中幼稚B细胞、浆细胞、活化树突细胞和中性粒细胞比例高于NAC样本。EVI2B表达量与M2巨噬细胞、M1巨噬细胞、CD8 T细胞、浆细胞、辅助滤泡T细胞、M0巨噬细胞和中性粒细胞呈正相关;与静息树突细胞呈负相关。结论 AAA发病涉及多种免疫细胞和信号通路,EVI2B在AAA样本中表达显著增高,与多种免疫细胞具有相关性,可能成为AAA治疗的新靶点。

关键词:腹主动脉瘤;加权基因共表达网络分析;机器学习;生物标志物;免疫浸润

中图分类号:R654.4

文献标志码:A

Comprehensive analysis of single-cell transcriptomics and machine learning reveals potential biomarkers for abdominal aortic aneurysm

WANG Yutao^{1,2}, SUN Yan³

(1. The First Clinical College, Shandong University of Traditional Chinese Medicine, Jinan 250355, Shandong, China;

2. Department of Peripheral Vascular Disease, Jinan Municipal Hospital of Traditional Chinese Medicine, Jinan 250012,

Shandong, China; 3. Department of Vascular Surgery, Shandong Provincial Hospital Affiliated to Shandong

First Medical University, Jinan 250021, Shandong, China)

收稿日期:2024-02-19

基金项目:国家自然科学基金青年科学基金项目(82104860);山东省中医药科技发展计划(2019-0559);济南市卫生健康委员会科技计划项目(2019-1-23);山东省医药卫生科技发展计划(2018WS478);济南市临床医学科技创新计划(202134013)

通信作者:孙岩. E-mail:sdxueguan@163.com

Abstract: Objective To screen for potential biomarkers of abdominal aortic aneurysm (AAA) using single-cell RNA (scRNA) data analysis, weighted gene co-expression network analysis (WGCNA), machine learning, and immune infiltration analysis. **Methods** The scRNA sequencing data containing AAA and normal aorta control (NAC) in the gene expression database were downloaded and processed by data quality control, dimensionality reduction, differential analysis, and cell type annotation. Chronological analysis was proposed to screen for the earliest differentiated cell types during AAA genesis, and to screen for differentially expressed genes (DEGs). High dimensional WGCNA (hdWGCNA) was performed to identify AAA-related gene modules, and enrichment analysis was conducted. Conventional transcriptome sequencing data containing AAA and NAC was downloaded for differential analysis and WGCNA. DEGs of scRNA samples, DEGs of conventional transcriptomes and WGCNA results were integrated to screen genes associated with AAA lesions. Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) signaling pathway enrichment analysis were carried out. The potential biomarkers of AAA were screened using the least absolute shrinkage and selection operator (LASSO), support vector machine recursive feature elimination (SVM-RFE), and random forest (RF) machine learning methods. The immune infiltration analysis was performed. **Results** The results of scRNA data analysis showed that endothelial cell was the earliest cell type to differentiate during AAA development, and a total of 853 scRNA DEGs were obtained. hdWGCNA identified 2 gene modules associated with AAA, which were significantly enriched in the signaling pathways of T helper 17 cell differentiation, and T helper 1 and 2 cell differentiation. Conventional transcriptome analysis yielded a total of 162 DEGs. Integration yielded 17 AAA-associated genes, which significantly enriched in signaling pathways such as chemokines, T helper 17 cell differentiation, and T helper 1 and 2 cell differentiation. The machine learning algorithm identified a potential biomarker for AAA, ecotropic viral integration site 2B (EVI2B). The expression of EVI2B was higher in AAA samples than in NAC samples. The immune infiltration results showed that the proportions of naive B cells, plasma cells, activated dendritic cells and neutrophils were higher in AAA samples than in NAC samples. EVI2B was positively correlated with M2 macrophages, M1 macrophages, CD8 T cells, plasma cells, helper follicular T cells, M0 macrophages, and neutrophils; and it was negatively correlated with resting dendritic cells. **Conclusion** AAA pathogenesis involves a variety of immune cells and signaling pathways, and EVI2B expression is significantly increased in AAA samples, correlating with a variety of immune cells, which may be a new target for AAA treatment.

Key words: Abdominal aortic aneurysm; Weighted gene co-expression network analysis; Machine learning; Biomarkers; Immune infiltration

腹主动脉瘤 (abdominal aortic aneurysm, AAA) 是常见的主动脉疾病, 腹主动脉最大直径超过 30 mm, 即可确诊^[1]。AAA 发病存在显著的性别差异, 据统计 60 岁以上男性的 AAA 患病率约为 4%~8%, 60 岁以上女性的 AAA 患病率约为 0.5%~1.5%^[2]。大部分 AAA 瘤体直径可随病程进展而增大, 瘤体破裂风险随直径的增大而上升^[3]。据统计, 直径 33~55 mm、56~60 mm 和 61~70 mm AAA 的 3 年年累积破裂发生率分别为 0.6%、2.2% 和 6.0%^[4-5]。破裂 AAA 的死亡率高达 80%^[6]。AAA 的病理生理机制复杂, 涉及脂质代谢、血管发育和重构、细胞外基质失调和炎症反应等^[7], 但其内在机制仍未完全明确。目前尚无有效预测 AAA 发生和发展的方法和生物标志物^[8-9], 因此, 阐明 AAA 发生的潜在机制具有重要意义。

单细胞转录组测序 (single-cell RNA sequencing, scRNA) 技术可通过高分辨率的组学信息, 更为精确地挖掘 AAA 的潜在机制^[10]。本研究通过分

析 AAA 样本的单细胞转录组测序数据, 利用加权基因共表达网络分析 (weighted gene co-expression network analysis, WGCNA)、最小绝对收缩和选择算子 (least absolute shrinkage and selection operator, LASSO)、支持向量机递归特征消除 (support vector machine recursive feature elimination, SVM-RFE) 和随机森林 (random forest, RF) 等机器学习方法筛选 AAA 的潜在生物标志物, 探讨标志物和浸润性免疫细胞之间的关系, 以期为 AAA 发病机制研究提供参考。

1 资料与方法

1.1 数据获取和质量控制

1.1.1 基因表达数据获取

从基因表达数据库 (GEO, <http://www.ncbi.nlm.nih.gov/geo>) 获得了 AAA 基因表达芯片数据集 GSE7084^[11]、GSE47472^[12]、GSE57691^[13]、GSE183464^[14] 和

2个 AAA scRNA 数据集(GSE226492 和 GSE166676^[15])。6个数据集中均包含 AAA 样本和作为对照的正常

主动脉样本(normal aorta control, NAC),数据集信息见表 1。

表 1 数据集信息
Table 1 Information of datasets

数据集	类型	样本(例)		测序平台
		AAA	NAC	
GSE7084	RNA	6	7	GPL2507
GSE47472	RNA	14	8	GPL10558
GSE57691	RNA	49	10	GPL10558
GSE183464	RNA	7	7	GPL20301
GSE226492	scRNA	3	3	GPL24676
GSE166676	scRNA	4	2	GPL24676

1.1.2 RNA 数据和 scRNA 数据质量控制

采用“limma”R 包对 RNA 数据中 AAA 组和 NAC 组样本测序数据进行标准化处理。利用“Seurat”R 包的“Read 10X”分别读取 GSE226492 和 GSE166676 数据集中每个样本的数据,根据类别,将样本分为 AAA 组和 NAC 组。筛选包含 200~5 000 个基因的细胞,以线粒体基因表达比例 <20%、红细胞基因表达比例 <5% 为截断值,对 scRNA 数据进一步过滤。使用“harmony”R 包去除两个数据集中的测序数据的批次效应。

1.2 方法

1.2.1 RNA 数据和 scRNA 数据差异表达基因(differential expressed genes, DEGs)筛选

以 $|\log_2 FC| > 1$ 、矫正后 $P < 0.05$ 为条件,筛选得到 DEGs,使用火山图和热图进行可视化。筛选 scRNA 数据前 2 000 个高变基因进行主成分分析,将主成分的数量调整为 15 个,对细胞进行分群,使用统一流形逼近与投影(uniform manifold approximation and projection, UMAP)图展示。使用“Seurat”R 包中的“FindAllmarkers”功能,筛选每个细胞分群中的 DEGs。

1.2.2 scRNA 细胞类型注释

登录 CellMarker 2.0 数据库(<http://117.50.127.228/CellMarker/index.html>),筛选各细胞分群 DEGs 对应的细胞类型,手动注释 scRNA 数据的细胞类型,计算 AAA 与 NAC 样本细胞类型的差异。

1.2.3 scRNA 拟时序分析

细胞类型注释后,使用 Monocle2 进行拟时序分析。使用“plot cell trajectory”函数来可视化细胞的分化轨迹,筛选最早分化的细胞类型,使用“FindMarkers”函数识别该类型细胞 AAA 与 NAC 样本间的 DEGs。

1.2.4 scRNA 高维 WGCNA 分析(high dimensional WGCNA, hdWGCNA)

使用“hdWGCNA”包对拟时序分析中最早分化

的细胞类型进行分析,去除批次效应,构建基因共表达网络,识别与 AAA 和拟时序分析中最早分化的细胞类型高度相关基因的模块,并计算模块特征基因连通性(eigengene-based connectivity, kME)值,提取高度相关基因模块中的基因进行功能富集分析。

1.2.5 RNA 数据 WGCNA 分析

利用“WGCNA”R 包构建共表达网络,筛选最佳软阈值,计算拓扑重叠矩阵(topological overlap matrix, TOM)。使用“hclust”函数进行层次聚类,根据拓扑重叠异度(1-TOM)对基因进行模块化处理,利用“Dynamic Tree Cut”和“Module Membership”函数筛选与 AAA 病变相关的基因模块。

1.2.6 AAA 相关基因筛选

将 scRNA 样本 DEGs、RNA 数据 DEGs 和 WGCNA 筛选获得的与 AAA 病变相关的基因模块取交集,获得 AAA 相关基因。使用 Monocle2 对相关基因进行拟时序分析,使用“clusterProfiler”R 包对相关基因进行基因本体论(gene ontology, GO)和京都基因和基因组百科全书(Kyoto Encyclopedia of Genes and Genomes, KEGG)信号通路富集分析。

1.2.7 机器学习筛选 AAA 生物标志物

使用“randomForest”R 包、“glmnet”R 包和“e1071”R 包对 RNA 数据中 AAA 组和 NAC 组样本测序数据进行 RF、LASSO 和 SVM-RFE 分析,将三种模型筛选获得的基因取交集,筛选 AAA 生物标志物。比较 AAA 生物标志物在 RNA 数据和 scRNA 数据中的表达水平。

1.2.8 免疫浸润分析

使用反卷积算法(CIBERSORT)分析 RNA 数据中 AAA 组和 NAC 组样本免疫细胞浸润水平,计算每种免疫细胞在两组样本中的比例以及不同免疫细胞之间的相关性。

1.2.9 免疫细胞比例与 AAA 生物标志物表达的相关性分析

分析免疫细胞比例与 AAA 生物标志物表达之间的相关性,使用“ggplot2”R 包可视化。使用 R 4.1.2 进行 Spearman 秩相关检验,分析免疫细胞与 AAA 生物标志物之间的关系, $P < 0.05$ 为差异有统计学意义。

2 结果

2.1 scRNA 质量控制与降维聚类

在 scRNA 数据集上执行质量控制过程。去除死亡细胞、双细胞数据,控制线粒体和红细胞基因比例,保证研究中使用的细胞样本的质量(图 1A)。筛选高变基因并进行降维处理,将细胞分为 12 个群

(图 1B~D)。

2.2 细胞注释

筛选每个细胞分群中的 DEGs,对 scRNA 数据的细胞类型进行手动注释。分群 0 注释为成纤维细胞,分群 1 注释为巨噬细胞,分群 2、3 注释为 T 细胞,分群 4 注释为内皮细胞,分群 5 注释为平滑肌细胞,分群 6 注释为 B 细胞,分群 7 注释为肥大细胞,分群 8 注释为单核细胞,分群 9 注释为脂肪细胞,分群 10 注释为树突状细胞,分群 11 注释为施万细胞,分群 12 注释为自然杀伤 T 细胞(图 2A~B)。AAA 样本 T 细胞、单核细胞、树突状细胞比例高于 NAC 样本,NAC 样本中平滑肌细胞、施万细胞、自然杀伤 T 细胞、肥大细胞、巨噬细胞、成纤维细胞、内皮细胞、B 细胞和脂肪细胞比例高于 AAA 样本(图 2C)。

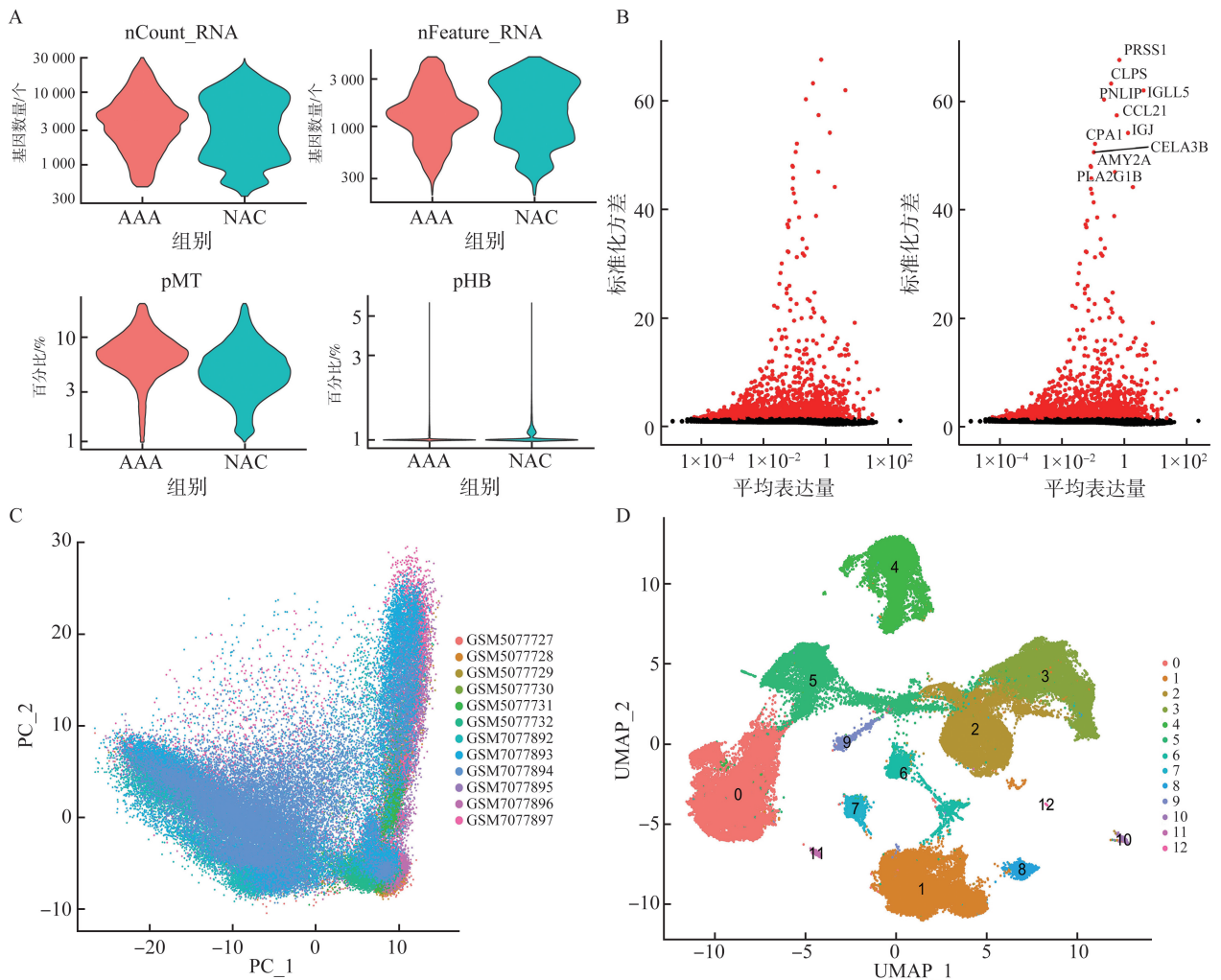


图 1 单细胞质量控制与降维聚类

A: 调整线粒体和红细胞基因比例,保证细胞样本质量;B: 2 000 个高变基因用红色表示,标注前 10 个基因名称;C: 降维并去批次效应后显示各样本的数据重叠良好,具有可比性;D: 细胞分群后 UMAP 图。

Figure 1 Single-cell quality control and dimensionality reduction clustering

A: Adjusting mitochondrial and red blood cell gene proportions to ensure cell sample quality; B: A total of 2,000 highly variable genes represented in red, with labels for the top 10 gene names; C: Dimensionality reduction and batch effect removal show well-overlapping data among samples, ensuring comparability; D: UMAP plot after cell clustering.

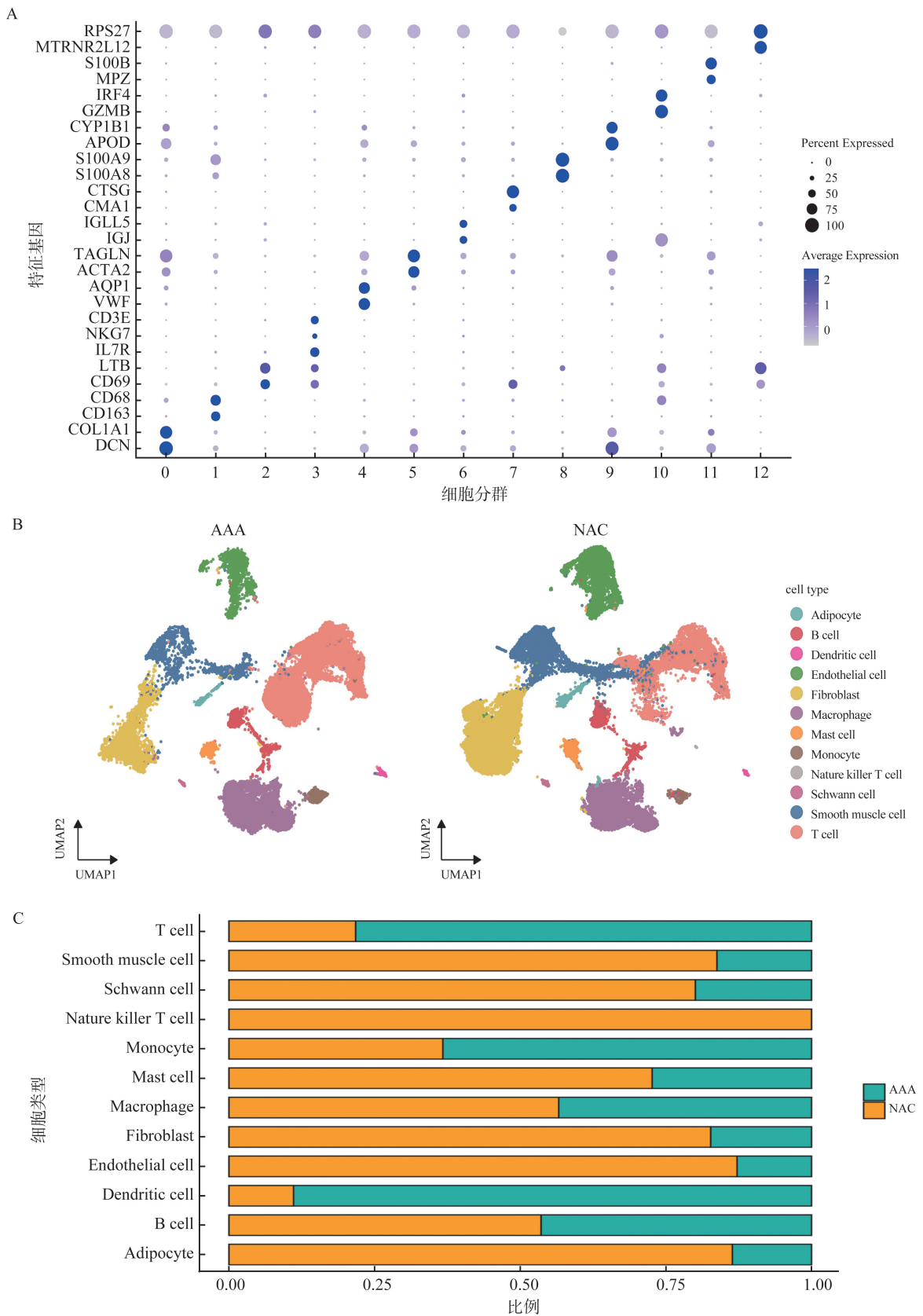


图 2 细胞注释
 A: 筛选各细胞分群的标记基因; B: 细胞注释后 AAA 样本与 NAC 样本的 UMAP 图; C: AAA 样本与 NAC 样本不同细胞比例柱状图。

Figure 2 Single-cell annotation

A: Selection of marker genes for each cell cluster; B: UMAP plot showing cell annotation for AAA and NAC samples; C: Bar chart depicting proportions of different cell types between AAA and NAC samples.

2.3 拟时序分析

对细胞分化轨迹进行拟时序分析,结果显示,随着时间的推移,细胞分化轨迹为从左向右(图 3A),内皮细胞是最早分化的细胞类型(图 3B),AAA 样本内皮细胞分化早于 NAC 样本内

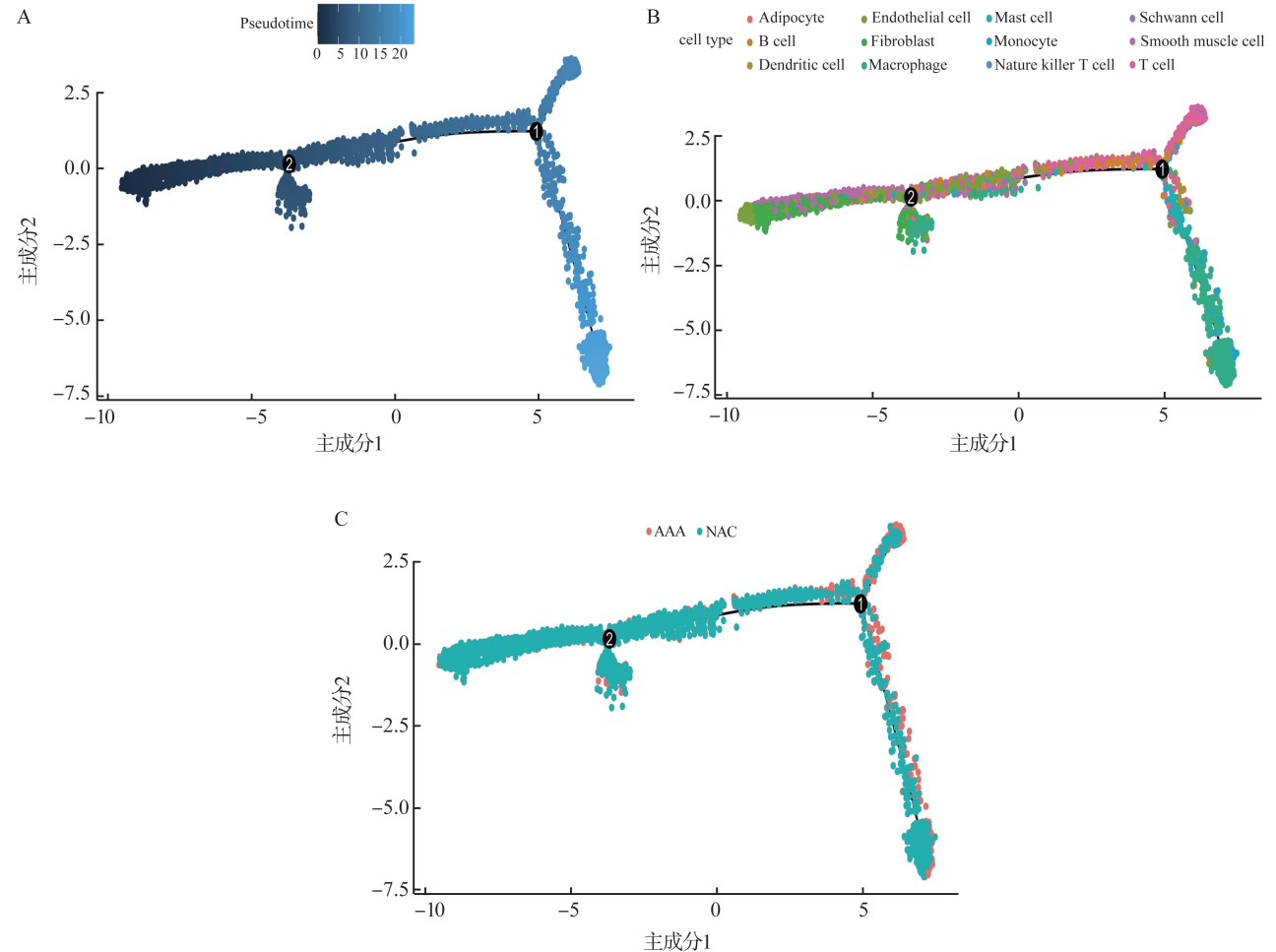


图 3 scRNA 拟时序分析

A: 细胞分化的时间差异,深蓝色代表分化的早期阶段,浅蓝色代表分化的晚期阶段;B: 腹主动脉所有细胞的分化轨迹;C: AAA 与 NAC 样本细胞的分化差异;①②代表细胞分化的时间节点。

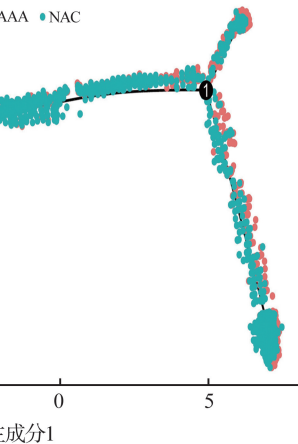
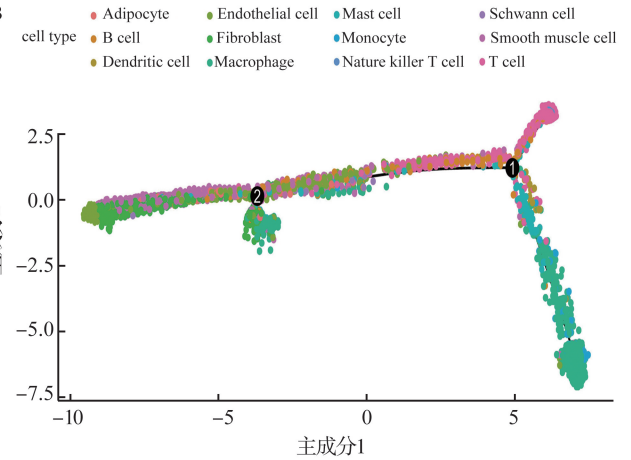
Figure 3 ScRNA pseudo-temporal analysis

A: Time differences in cell differentiation, where deep blue represents early stages of differentiation and light blue represents late stages; B: Differentiation trajectories for all cells of abdominal aorta; C: Cell differentiation between AAA and NAC samples; ① and ② represent the time point of cell differentiation.

2.4 hdWGCNA

利用 hdWGCNA 挖掘内皮细胞的特性和功能。在对中位连通性、平均连通性和无标度拓扑模型拟合进行加权后,选择软阈值为 9(图 4A),并生成 13 个模块(图 4B)。在这些模块中,EC-M6 和 EC-M11 模块中的基因在内皮细胞中特异性表达,且与内皮细胞呈正相关(图 4C~D)。通过 kME 值筛选核心基因(图 4E),富集分析结果显示,EC-M6 和 EC-M11 模块基因显著富集于辅助性 T 细胞 17 细胞分化、辅助性 T 细胞 1 和 2 细胞

分化等信号通路,以及细胞质翻译等生物学过程、细胞质核糖体等细胞组分和核糖体结构成分等细胞功能上(图 4F~G)。



分化等信号通路,以及细胞质翻译等生物学过程、细胞质核糖体等细胞组分和核糖体结构成分等细胞功能上(图 4F~G)。

2.5 RNA 数据 DEGs 筛选

使用“limma”R 包对 RNA 数据中 AAA 组和 NAC 组样本测序数据进行标准化处理并筛选 DEGs(图 5A~B),共获得 4 862 个表达量大于 0 的基因,筛选获得 162 个 DEGs,其中差异表达上调基因 55 个,差异表达下调基因 107 个。绘制 DEGs 火山图和热图(图 5C~D)。

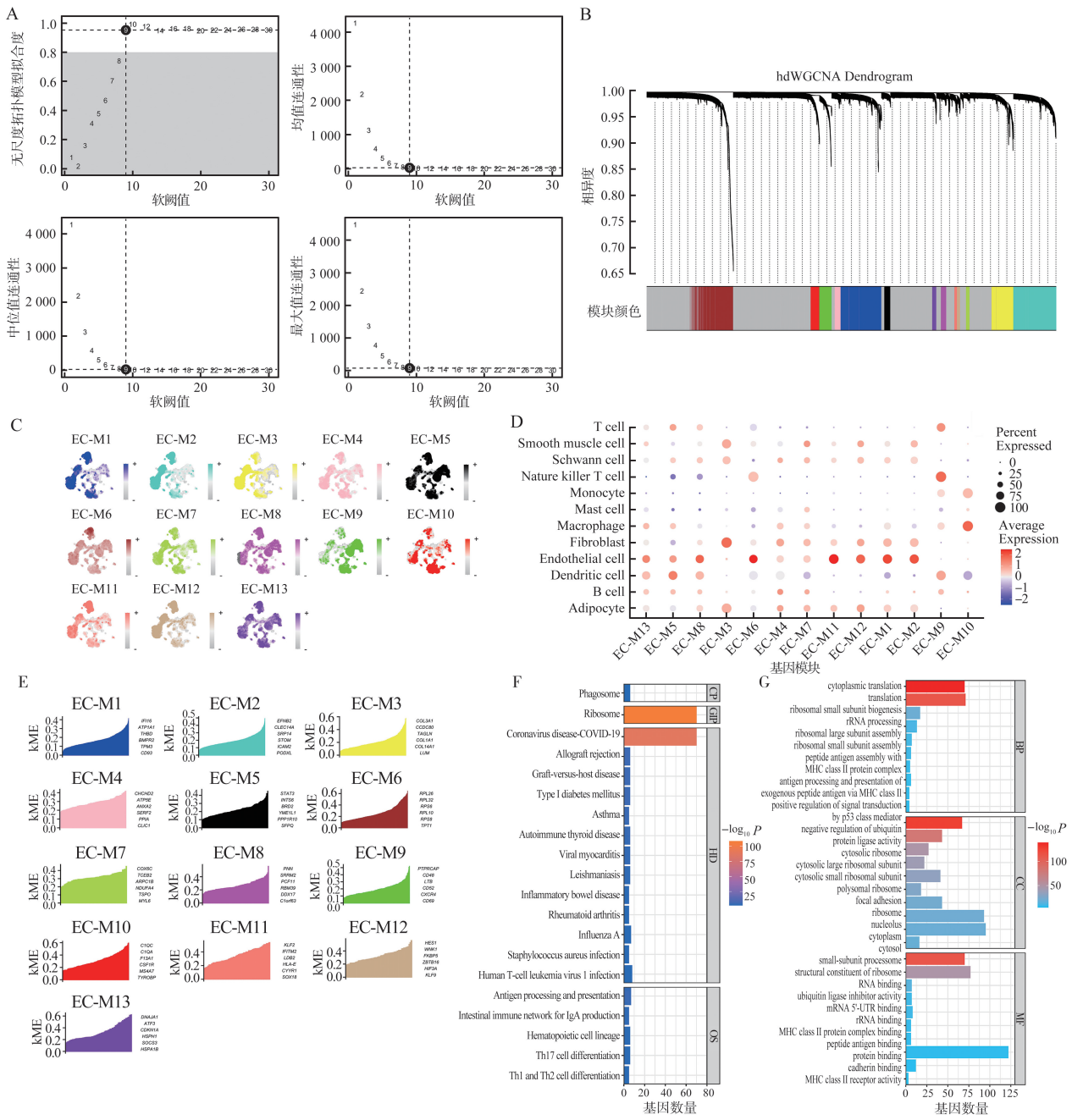


图4 内皮细胞 hdWGCNA

A: 左上角的面板描绘了选择 ≥ 0.9 的无尺度拓扑模型拟合的最佳软阈值为9; 右上角、左下角及右下角三个面板分别显示了选择不同最小软阈值时拓扑网络的平均值、中位数和最大连通性, 当最佳软阈值为9时, 拓扑网络的平均连通性最稳定; B: hdWGCNA 树状图, 筛选出13个模块; C: 各模块与内皮细胞相关性评分, 颜色越深, 代表相关性越高; D: 各模块与不同细胞类型的相关性, 红色代表正相关, 蓝色代表负相关, 颜色越深, 代表相关性越高; E: 根据 KME 值对每个模块中的核心基因进行鉴定和排序; F: EC-M6 和 EC-M11 模块基因 KEGG 信号通路富集分析; G: EC-M6 和 EC-M11 模块核心 GO 富集分析。

Figure 4 hdWGCNA of endothelial cell

A: Top left panel depicted the soft power threshold ≥ 9 , for choosing a scale-free topology model fit greater than or equal to 0.9. The bottom left panel, top right panel, and bottom right panel showed the mean, median, and max connectivity of the topological network respectively when different minimum soft thresholds were chosen, reflecting the connectivity of the network. The average connectivity of the topological network was most stable at the lowest soft threshold equals 9; B: Thirteen modules were identified as shown in the hdWGCNA dendrogram; C: The correlation score between the modules and endothelial cell, and the darker the color, the higher the correlation; D: The correlation of the modules with different cell types, where red indicates positive correlation, blue indicates negative correlation and the darker the color, the higher the correlation; E: Hub genes in each module were identified and ranked by kME; F: KEGG pathway enrichment analysis of EC-M6 module and EC-M11 module related genes; G: GO enrichment analysis of EC-M6 module and EC-M11 module related genes.

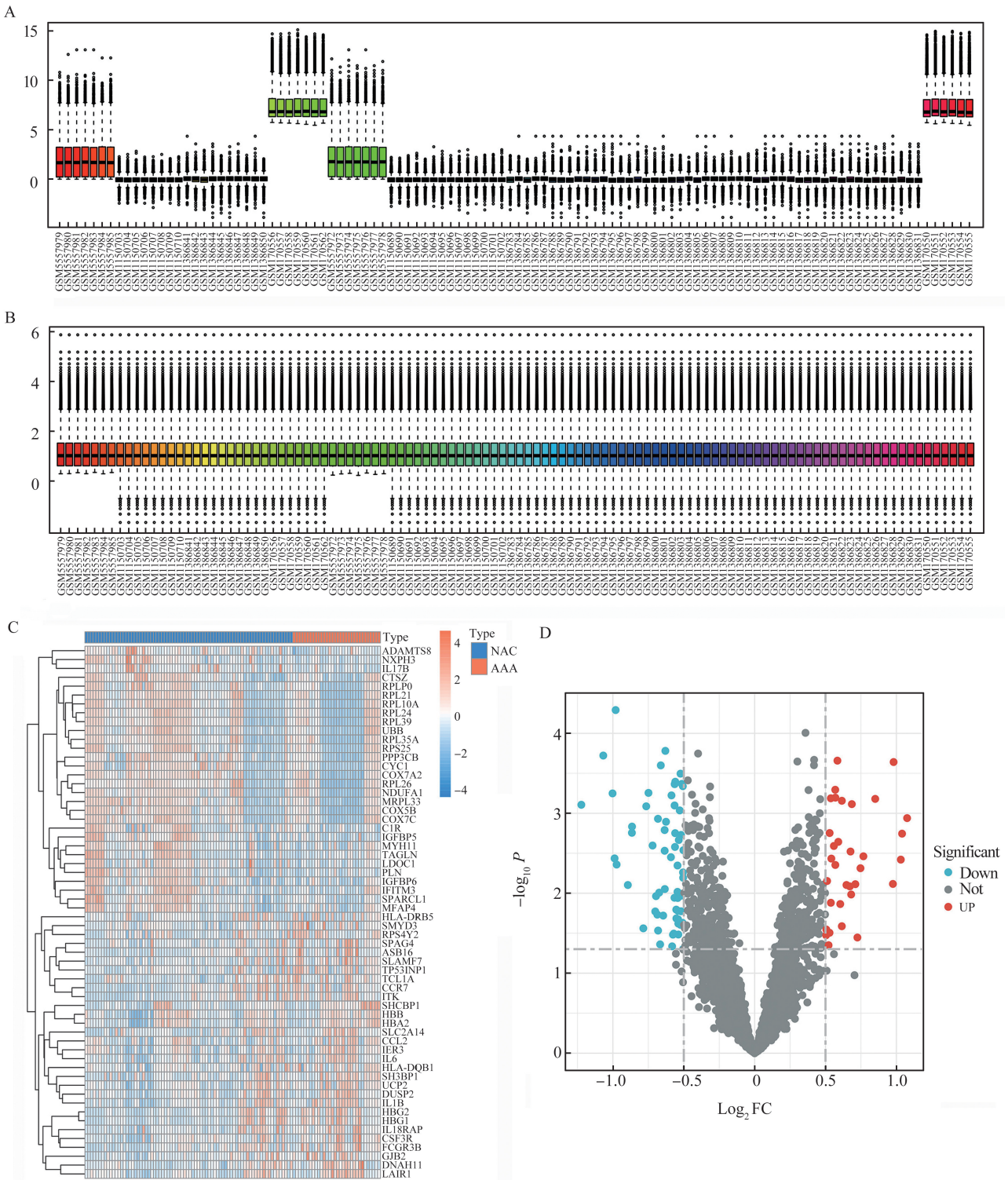


图5 RNA数据DEGs筛选

A: 标准化处理前的样本表达量柱状图; B: 标准化处理后的样本表达量柱状图; C: DEGs 热图; D: DEGs 火山图。

Figure 5 DEGs screening in RNA data

A: Bar chart of sample expression levels before normalization; B: Bar chart of sample expression levels after normalization; C: Heatmap of DEGs. D: Volcano plot of DEGs.

2.6 WGCNA

使用“WGCNA”包的“pickSoftThreshold”函数对4862个表达量大于0的基因进行筛选,将软阈值设为29,建立无尺度网络,将阈值设为

0.25,最小模块基因数设为50,共聚类出7个模块(图6A)。模块-性状关联分析显示,黑色、蓝色和紫红色模块与AAA显著相关(图6B),包括151个基因。

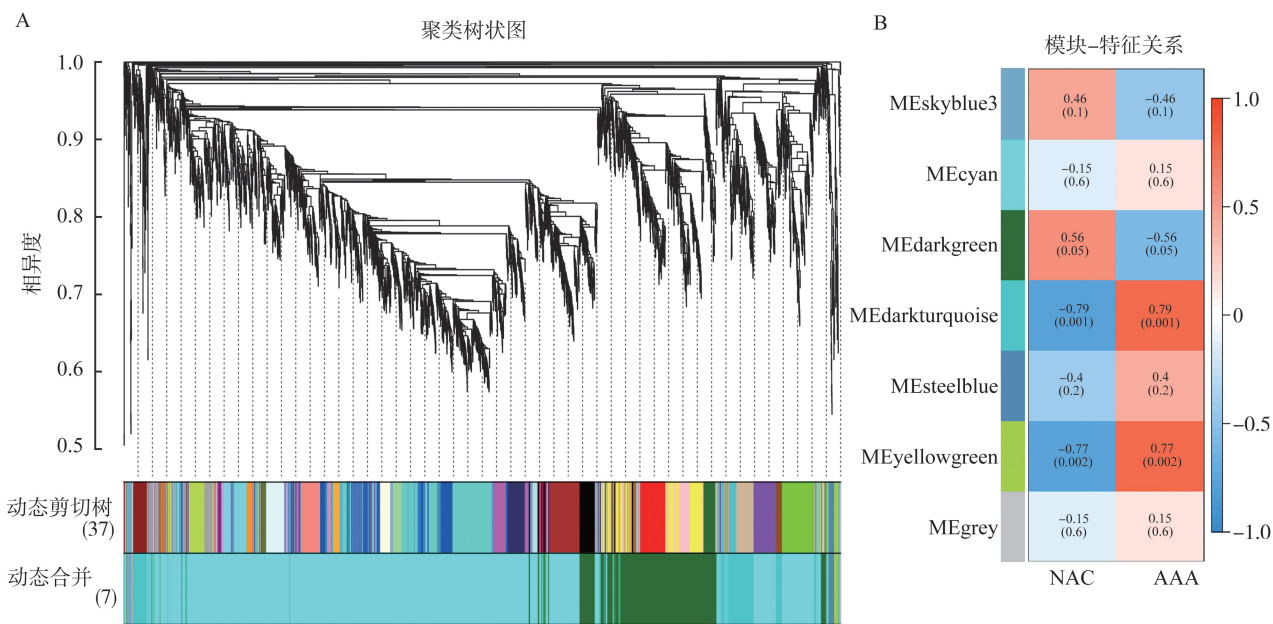


图6 RNA数据加权基因共表达网络分析

A:数据集模块聚类树状图;B:基因模块和临床性状相关性热图。

Figure 6 WGCNA in the RNA data

A: Hierarchical cluster analysis; B: Heatmap of the correlation between the module eigengenes and clinical traits.

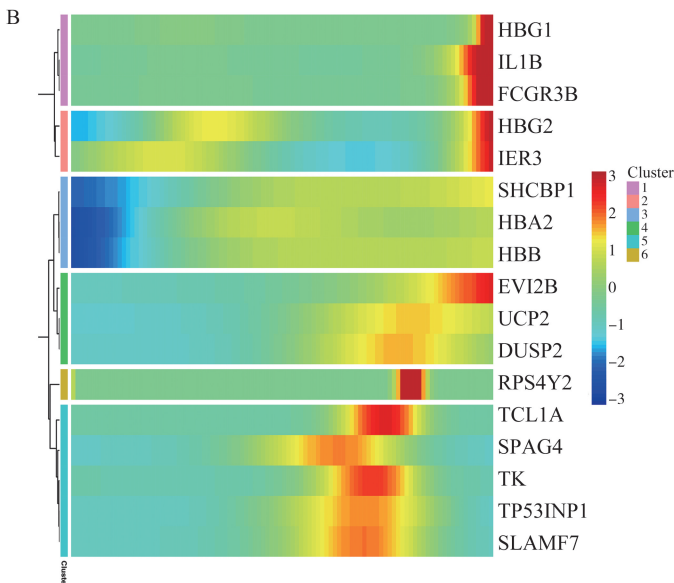
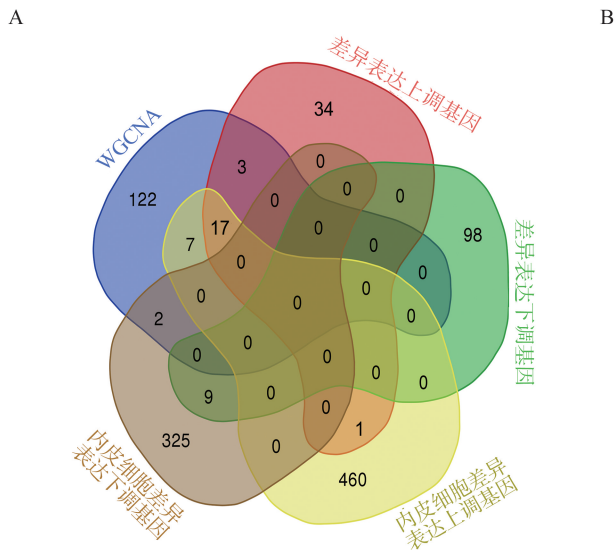
2.7 AAA相关基因筛选和富集分析

将scRNA样本DEGs、RNA数据DEGs和WGCNA筛选获得的与AAA病变相关的基因模块取交集,获得17个AAA相关基因(图7A)。对AAA相关基因进行拟时序分析,结果显示,随着疾病的进展,AAA相关基因均呈表达上调趋势(图7B)。富集分析结果显示,AAA相关基因显著富集于细胞黏附分子、辅助性T细胞17细胞分化、辅助性T细胞1和2细胞分化等信号通路,以及T细胞活化负调控等生物学过程、细胞外泌体等细胞组分和MHC II类蛋白复合物结合等细胞功能

(图7C~D)。

2.8 AAA生物标志物识别

使用LASSO回归从AAA相关基因中识别出8个关键的生物标志物(图8A),使用RF算法识别出7个关键生物标志物(图8B),使用SVM-RFE法识别出5个关键生物标志物(图8C),三种方法筛选出1个关键生物标志物EVI2B(图8D)。提取RNA数据和scRNA数据中EVI2B的表达量并进行分析,结果显示,RNA数据和scRNA数据中AAA样本EVI2B表达量均显著高于NAC样本(图9)。



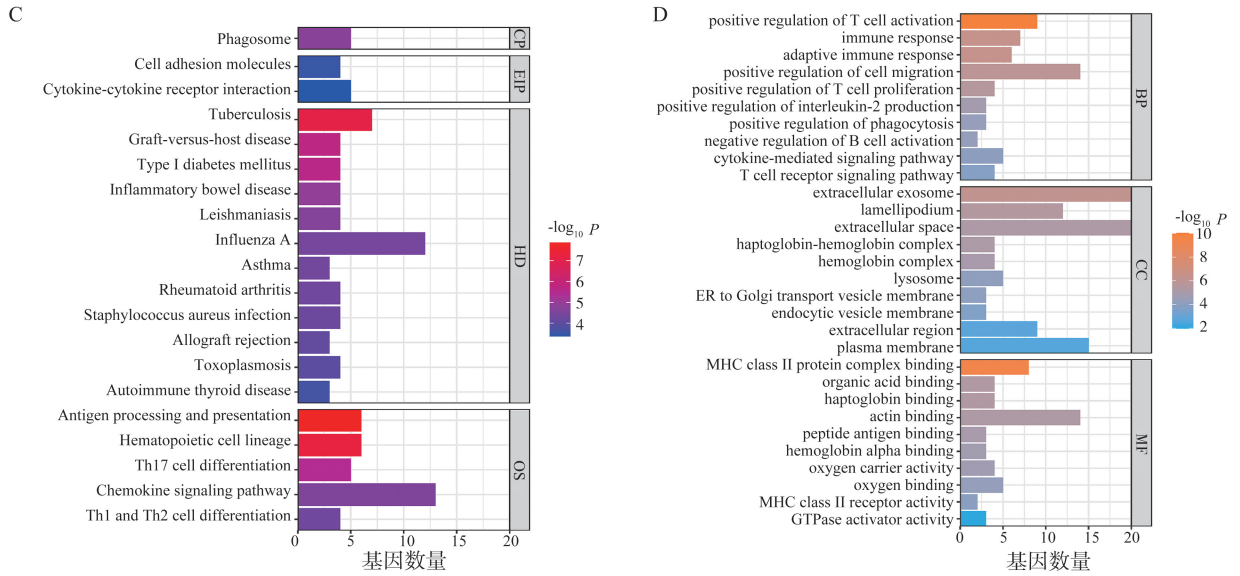


图7 AAA 相关基因筛选和富集分析

A:AAA 相关基因筛选韦恩图;B:AAA 相关基因 scRNA 拟时序分析热图;C:AAA 相关基因 KEGG 信号通路富集分析; D:AAA 相关基因 GO 富集分析。

Figure 7 Screening and enrichment analysis of AAA-related genes

A: Venn diagram of screened AAA-related genes; B: ScRNA pseudo-temporal analysis heatmap of AAA-related genes; C: KEGG pathway enrichment analysis of AAA-related genes; D: GO enrichment analysis of AAA-related genes.

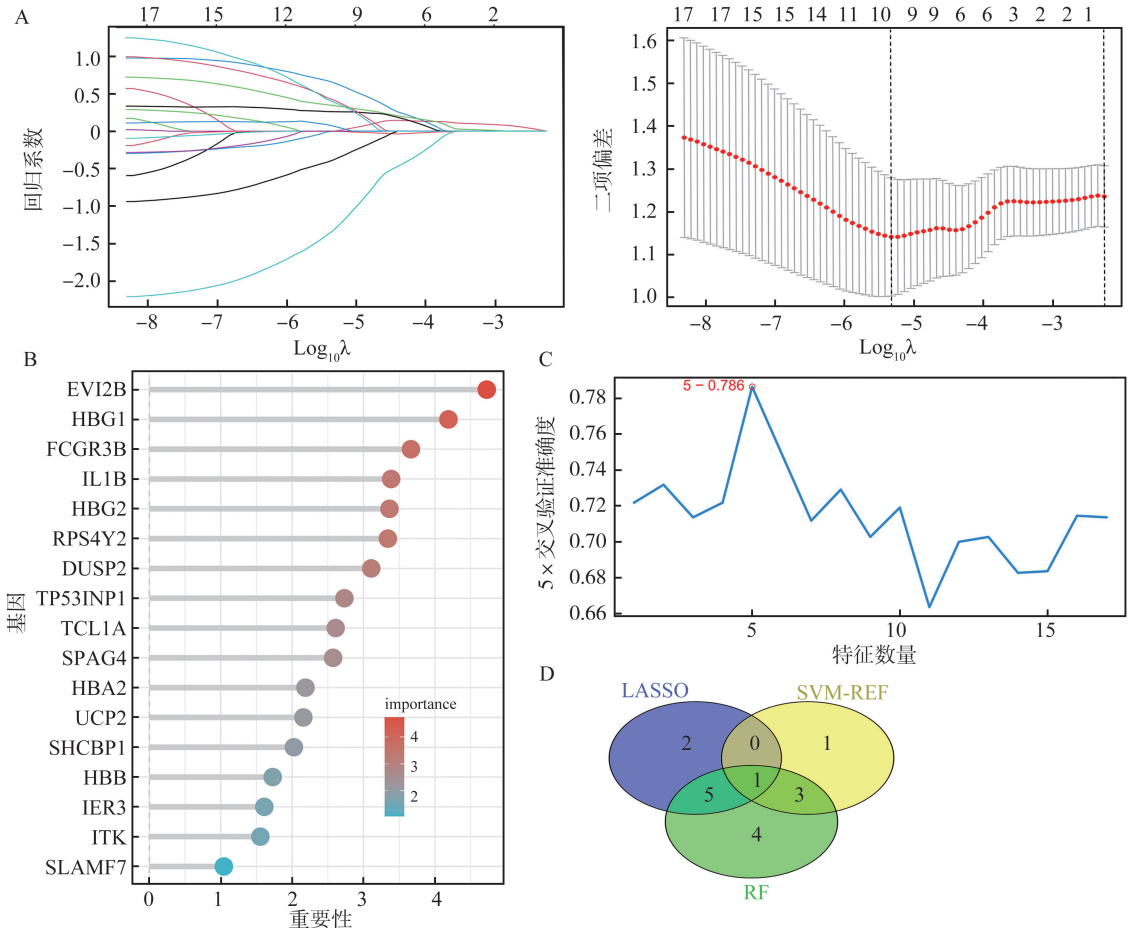


图8 AAA 生物标志物识别

A:LASSO 回归识别 AAA 生物标志物;B:RF 识别 AAA 生物标志物;C:SVM-RFE 识别 AAA 生物标志物;D:三种机器学习识别 AAA 生物标志物的韦恩图。

Figure 8 Identification of AAA biomarkers

A: LASSO regression identified AAA biomarkers; B: RF identified AAA biomarkers; C: SVM-RFE identified AAA biomarkers; D: Venn diagram of AAA biomarkers identified by three machine learning methods.

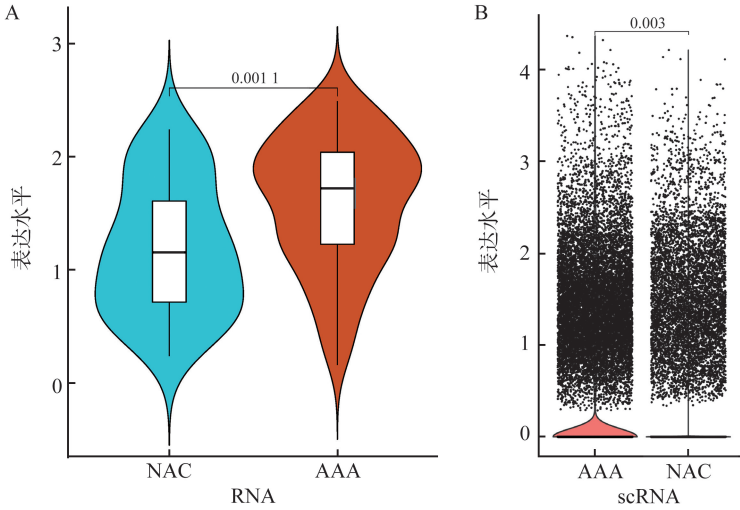


图9 EVI2B 表达量小提琴图

A: RNA 数据中 EVI2B 表达差异; B: scRNA 数据中 EVI2B 表达差异。

Figure 9 Violin plots of EVI2B expression

A: Differential expression of EVI2B in RNA data; B: Differential expression of EVI2B in scRNA data.

2.9 免疫浸润分析

使用 CIBERSORT 算法对 RNA 数据 AAA 和 NAC 样本进行免疫浸润分析 (图 10A)。22 个免疫细胞相关性热图结果显示, 幼稚 T 细胞与调节性 T

细胞, 树突细胞与嗜酸性粒细胞呈正相关 (图 10B)。AAA 样本中幼稚 B 细胞、浆细胞、活化树突细胞和中性粒细胞比例高于 NAC 样本 (图 10C)。

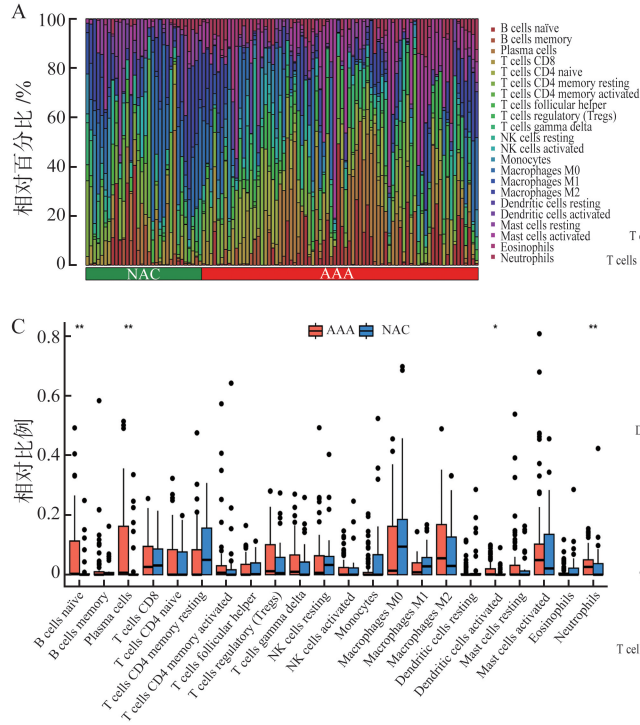


图10 免疫浸润分析

A: AAA 和 NAC 样本 22 个免疫细胞亚群比较热图; B: AAA 和 NAC 样本中 22 种不同免疫细胞之间相关性的热图, 红色表示正相关, 蓝色表示负相关, 相关性越强, 颜色越深; C: AAA 和 NAC 样本中 22 种免疫细胞的比例差异的柱状图。

Figure 10 Immune infiltration analysis

A: Heatmap comparing 22 immune cell subtypes between AAA and NAC samples; B: Heatmap shows the correlation between 22 different immune cell types in AAA and NAC samples, where red indicates positive correlation, blue indicates negative correlation, and the stronger the correlation, the darker the color; C: Bar chart shows the proportional differences of 22 immune cell types between AAA and NAC samples.

2.10 EVI2B 表达与 AAA 样本免疫细胞比例的相关性分析

EVI2B 与 M2 巨噬细胞 ($r = 0.59, P < 0.001$)、M1 巨噬细胞 ($r = 0.48, P < 0.001$)、CD8 T 细胞 ($r = 0.43, P < 0.001$)、浆细胞 ($r = 0.37, P < 0.001$)、辅助

滤泡 T 细胞 ($r = 0.36, P < 0.001$)、M0 巨噬细胞 ($r = 0.33, P < 0.001$) 和中性粒细胞 ($r = 0.26, P < 0.023$) 呈正相关;与静息树突细胞 ($r = -0.26, P = 0.026$) 呈负相关(图 11)。

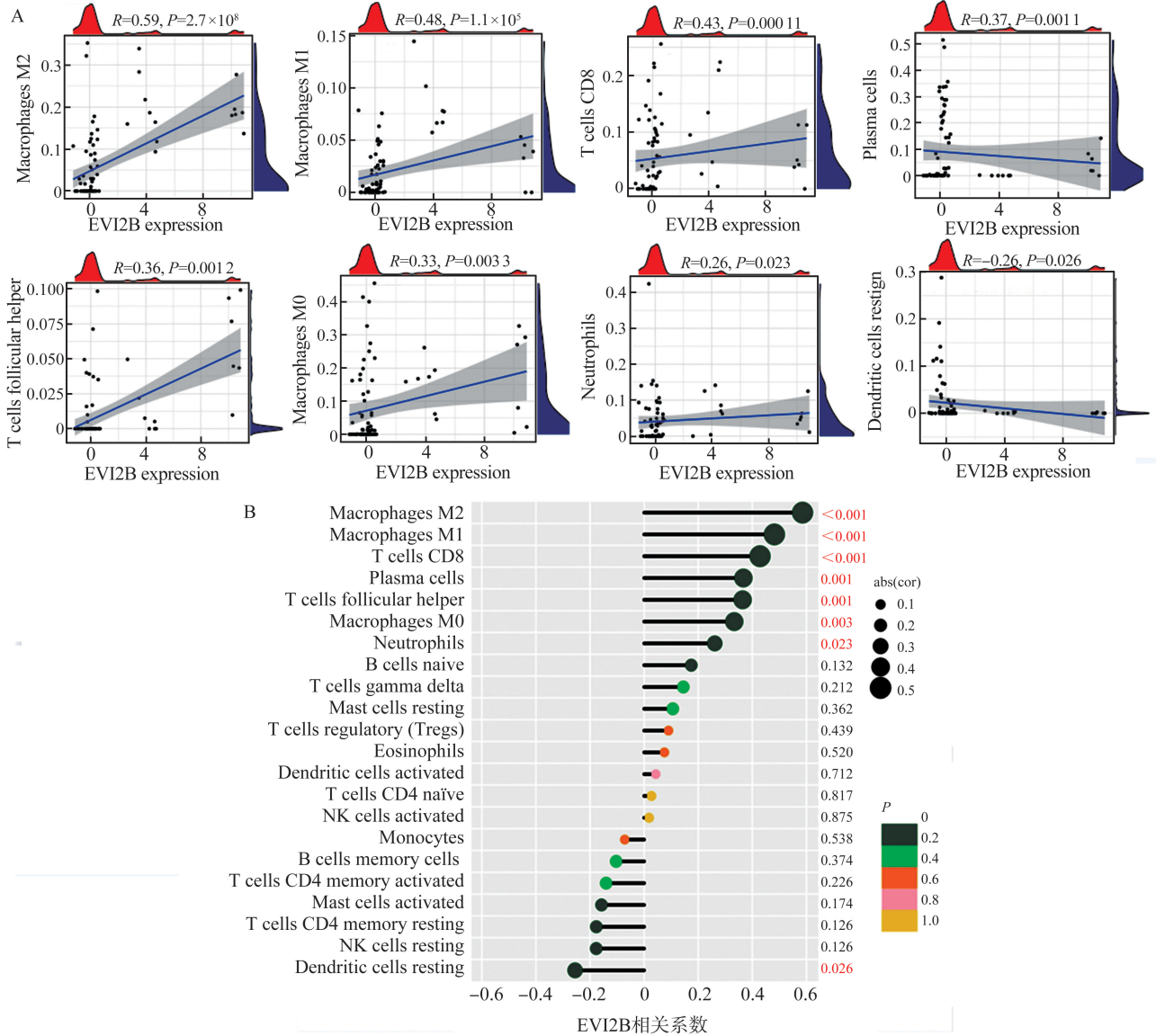


图 11 EVI2B 表达量与免疫细胞的相关性分析

A: EVI2B 表达量与免疫细胞比例间的相关性散点图; B: EVI2B 表达量与免疫细胞比例间的相关性棒棒糖图。

Figure 11 The correlation analysis between EVI2B expression and immune cells proportion

A: Scatter plot of correlation between EVI2B expression and proportion of immune cells; B: Lollipop plot of correlation between EVI2B expression and proportion of immune cells.

3 讨论

本研究通过拟时序分析发现,内皮细胞是 AAA 病变过程中最早分化的细胞类型,提示内皮细胞在 AAA 发生和发展过程中发挥重要作用。附壁血栓和内皮细胞损伤是 AAA 的主要特征^[16]。内皮细胞在调节血管和血流稳态中发挥重要作用,还参与调

节血管张力、血管生成、平滑肌细胞增殖和炎症反应^[17]。内皮细胞的内皮型一氧化氮合酶产生一氧化氮,刺激环鸟苷酸生成,引发一系列细胞内事件,导致平滑肌松弛,诱发 AAA 发病^[18]。稳定的腹主动脉内皮剪切应力可通过阻碍内皮细胞释放细胞因子和黏附分子等途径,减少炎症介质的生成和腹主动脉管壁重塑,降低腹主动脉外膜的炎症反应,抑制胶原纤维的降解,保护腹主动脉弹性纤维和平滑肌

细胞的功能损伤^[19]。Zhao等^[20]研究证实,敲除内皮细胞周期蛋白I可抑制内皮依赖性血管舒张和血流动力学反应,影响血管重塑,增加AAA发生风险。

单细胞hdWGCNA和AAA相关基因富集分析结果均显示AAA相关基因显著富集于辅助性T细胞17细胞分化、辅助性T细胞1和2细胞分化等信号通路上。研究发现,在AAA患者的主动脉组织中,辅助性T细胞17细胞分泌的白细胞介素-17明显增加,促进了局部动脉炎症反应^[21]。López-Sanz等^[22]研究发现,AAA模型小鼠腹主动脉巨噬细胞和B淋巴细胞含量增加,白细胞介素-17、趋化因子CCL2、肿瘤坏死因子- α 和基质金属蛋白酶2的表达上调。Lu等^[23]研究发现,骨髓来源抑制性细胞在AAA患者血液中异常聚集,且能促进辅助性T细胞17细胞分化,导致AAA患者血清白细胞介素-3、辅助性T细胞17表达升高,参与AAA发病。Zhu等^[24]研究发现,AAA模型小鼠腹主动脉辅助性T细胞1/辅助性T细胞2细胞因子、干扰素- γ 、肿瘤坏死因子- α 、白细胞介素-4表达上调。Wang等^[25]研究结果显示,AAA患者外周血CD4细胞辅助性T细胞1特异性转录因子和炎症小体表达上调,辅助性T细胞2特异性转录因子表达下调。

本研究通过对AAA样本的单细胞转录组和常规转录组测序数据分析,并借助机器学习方法,筛选获得了潜在的AAA生物标志物生态病毒整合位点2B(ecotropic viral integration site 2B, EVI2B)。EVI2B已被证实在骨髓分化和造血祖细胞功能中发挥重要作用,参与白血病的发生^[26]。研究表明,EVI2B在骨肉瘤样本中呈高表达,参与单核细胞与巨噬细胞的浸润,其机制可能与影响成骨细胞和破骨细胞的分化和功能相关^[27]。动脉粥样硬化病理过程中,血管平滑肌细胞可向间充质干细胞转化,转化为间充质干细胞的血管平滑肌细胞可进一步分化为成骨细胞,诱发血管钙化^[28]。破骨细胞与成骨细胞在骨发育、骨修复和钙盐代谢的转换中发挥着重要作用。破骨细胞分化参与动脉粥样硬化的病理过程^[29]。形态学研究证实,在动脉粥样硬化模型小鼠的钙化斑块和动脉中膜钙化病变中存在多核破骨细胞样细胞^[30]。血管钙化是AAA发生的主要机制之一^[31]。基于上述研究,结合本研究结果,我们推测,EVI2B参与AAA的发生、发展过程,可能与其介导成骨细胞和破骨细胞的分化,参与血管平滑肌细胞表型的改变和钙化有关。本研究结果显示,EVI2B与M2巨噬细胞、M1巨噬细胞、CD8T细胞、浆细胞、辅助滤泡T细胞、M0巨噬细胞和中性粒细胞呈

亚相关,与静息树突细胞呈负相关,提示其广泛参与免疫反应,EVI2B参与AAA病理过程的机制有待进一步研究。

综上所述,从GEO数据库中获取的AAA样本的单细胞和常规转录组测序数据分析结果显示,内皮细胞可能是AAA病理过程中的关键细胞类型,细胞黏附分子、辅助性T细胞17细胞分化、辅助性T细胞1和2细胞分化等信号通路参与AAA的发生发展过程,EVI2B在AAA样本中高表达,是潜在的AAA生物标志物。上述结果为AAA的发病机制和潜在治疗靶点提供了新的见解。

参考文献:

- [1] 张韬,郭伟.腹主动脉瘤诊断和治疗中国专家共识(2022版)[J].中国实用外科杂志,2022,42(4):380-387.
- [2] Yuan Z, Lu Y, Wei J, et al. Abdominal aortic aneurysm: roles of inflammatory cells[J]. Front Immunol, 2020, 11: 609161. doi:10.3389/fimmu.2020.609161.
- [3] 王颖,顾慧,于鑫鑫,等.71例腹主动脉瘤的基线CT特征与病变进展的相关性[J].山东大学学报(医学版),2022,60(10):62-67.
WANG Ying, GU Hui, YU Xinxin, et al. Correlation between baseline CT features and progression of abdominal aortic aneurysm in 71 cases[J]. Journal of Shandong University (Health Sciences), 2022, 60(10): 62-67.
- [4] Kindon AJ, McCombie AM, Frampton C, et al. Early relative growth rate of abdominal aortic aneurysms and future risk of rupture or repair[J]. Eur J Vasc Endovasc Surg, 2023, 66(6): 797-803.
- [5] Lancaster EM, Gologorsky R, Hull MM, et al. The natural history of large abdominal aortic aneurysms in patients without timely repair[J]. J Vasc Surg, 2022, 75(1): 109-117.
- [6] Leone N, Broda MA, Eiberg JP, et al. Systematic review and meta-analysis of the incidence of rupture, repair, and death of small and large abdominal aortic aneurysms under surveillance[J]. J Clin Med, 2023, 12(21): 6837.
- [7] Roychowdhury T, Klarin D, Levin MG, et al. Genome-wide association meta-analysis identifies risk loci for abdominal aortic aneurysm and highlights PCSK9 as a therapeutic target[J]. Nat Genet, 2023, 55(11): 1831-1842.
- [8] Zhang S, Gu H, Chang N, et al. Assessing abdominal aortic aneurysm progression by using perivascular adipose tissue attenuation on computed tomography angiography[J]. Korean J Radiol, 2023, 24(10): 974-982.
- [9] Hofmann A, Khorzom Y, Klimova A, et al. Associations of tissue and soluble LOX-1 with human abdominal aortic

- aneurysm [J]. *J Am Heart Assoc*, 2023, 12 (14): e027537. doi: 10.1161/JAHA.122.027537.
- [10] Davis FM, Tsoi LC, Melvin WJ, et al. Inhibition of macrophage histone demethylase JMJD3 protects against abdominal aortic aneurysms[J]. *J Exp Med*, 2021, 218 (6): e20201839. doi:10.1084/jem.20201839.
- [11] Pahl MC, Erdman R, Kuivaniemi H, et al. Transcriptional (ChIP-chip) analysis of ELF1, ETS2, RUNX1 and STAT5 in human abdominal aortic aneurysm[J]. *Int J Mol Sci*, 2015, 16(5): 11229-11258.
- [12] Biros E, Moran CS, Rush CM, et al. Differential gene expression in the proximal neck of human abdominal aortic aneurysm[J]. *Atherosclerosis*, 2014, 233 (1): 211-218.
- [13] Biros E, Gäbel G, Moran CS, et al. Differential gene expression in human abdominal aortic aneurysm and aortic occlusive disease [J]. *Oncotarget*, 2015, 6 (15): 12984-12996.
- [14] Zhang KJ, Yue JN, Yin L, et al. Comprehensive bioinformatics analysis revealed potential key genes and pathways underlying abdominal aortic aneurysm[J]. *Comput Struct Biotechnol J*, 2023, 21: 5423-5433. doi: 10.1016/j.csbj.2023.10.052.
- [15] Davis FM, Tsoi LC, Melvin WJ, et al. Inhibition of macrophage histone demethylase JMJD3 protects against abdominal aortic aneurysms[J]. *J Exp Med*, 2021, 218 (6): e20201839. doi:10.1084/jem.20201839.
- [16] Manenti A, Farinetti A, Manco G, et al. Intraluminal Thrombus and abdominal aortic aneurysm complications [J]. *Ann Vasc Surg*, 2022, 83: e11-e12.
- [17] Zheng DD, Liu J, Piao HL, et al. ROS-triggered endothelial cell death mechanisms: focus on pyroptosis, parthanatos, and ferroptosis[J]. *Front Immunol*, 2022, 13: 1039241. doi:10.3389/fimmu.2022.1039241.
- [18] DeRoo E, Stranz A, Yang H, et al. Endothelial dysfunction in the pathogenesis of abdominal aortic aneurysm [J]. *Biomolecules*. 2022,12(4): 509. doi: 10.3390/biom12040509.
- [19] Spartalis E, Spartalis M, Athanasiou A, et al. Endothelium in aortic aneurysm disease: new insights [J]. *Curr Med Chem*, 2020, 27(7): 1081-1088.
- [20] Zhao W, Yao MY, Zhang YY, et al. Endothelial cyclin I reduces vulnerability to angiotensin II-induced vascular remodeling and abdominal aortic aneurysm risk[J]. *Microvasc Res*, 2022, 142: 104348. doi: 10.1016/j.mvr.2022.104348.
- [21] Sharma AK, Lu GY, Jester A, et al. Experimental abdominal aortic aneurysm formation is mediated by IL-17 and attenuated by mesenchymal stem cell treatment[J]. *Circulation*, 2012, 126(Suppl 1): S38-S45.
- [22] López-Sanz L, Bernal S, Jiménez-Castilla L, et al. The presence of activating IgG Fc receptors in macrophages aggravates the development of experimental abdominal aortic aneurysm [J]. *Clin Investig Arterioscler*, 2023, 35(4): 185-194.
- [23] Lu L, Jin Y, Tong YH, et al. Myeloid-derived suppressor cells promote the formation of abdominal aortic aneurysms through the IL-3-ICOSL-ICOS axis [J]. *BBA Adv*, 2023, 4: 100103. doi: 10.1016/j.bbadv.2023.100103.
- [24] Zhu HL, Qu XQ, Zhang CS, et al. Interleukin-10 promotes proliferation of vascular smooth muscle cells by inhibiting inflammation in rabbit abdominal aortic aneurysm [J]. *Int J Clin Exp Pathol*, 2019, 12(4): 1260-1271.
- [25] Wang H, Wei GD, Cheng S, et al. Circulatory CD4-positive T-lymphocyte imbalance mediated by homocysteine-induced AIM2 and NLRP1 inflammasome upregulation and activation is associated with human abdominal aortic aneurysm[J]. *J Vasc Res*, 2020, 57(5): 276-290.
- [26] Zjablovskaja P, Kardosova M, Danek P, et al. EVI2B is a C/EBP α target gene required for granulocytic differentiation and functionality of hematopoietic progenitors [J]. *Cell Death Differ*, 2017, 24(4): 705-716.
- [27] Dai ZL, Liu ZQ, Yang R, et al. EVI2B is a prognostic biomarker and is correlated with monocyte and macrophage infiltration in osteosarcoma based on an integrative analysis[J]. *Biomolecules*, 2023, 13(2): 327.
- [28] Yap C, Mieremet A, de Vries CJM, et al. Six shades of vascular smooth muscle cells illuminated by KLF4 (Krüppel-like factor 4) [J]. *Arterioscler Thromb Vasc Biol*, 2021, 41(11): 2693-2707.
- [29] Gu JQ, Yang WW, Lin S, et al. Identification of co-expressed genes and immune infiltration features related to the progression of atherosclerosis [J]. *J Appl Genet*, 2024, 65(2): 331-339.
- [30] Wierer M, Prestel M, Schiller HB, et al. Compartment-resolved proteomic analysis of mouse aorta during atherosclerotic plaque formation reveals osteoclast-specific protein expression [J]. *Mol Cell Proteomics*, 2018, 17 (2): 321-334.
- [31] Petsophonsakul P, Furmanik M, Forsythe R, et al. Role of vascular smooth muscle cell phenotypic switching and calcification in aortic aneurysm formation [J]. *Arterioscler Thromb Vasc Biol*, 2019, 39(7): 1351-1368.