• 数据挖掘与循证医学 •

基于生物信息学与机器学习探索多囊卵巢综合征关键基因及临床验证与中药靶向预测

张红阳1*, 王钦鹏2, 王晓姗1, 王勇淅3

- 1. 赤峰学院附属医院 中医科, 内蒙古 赤峰 024005
- 2. 赤峰学院中医(蒙医)学院,内蒙古 赤峰 024000
- 3. 内蒙古医科大学中医学院,内蒙古 呼和浩特 010000

摘 要:目的 基于生物信息学与机器学习方法探讨多囊卵巢综合征(polycystic ovary syndrome, PCOS)关键基因,并在 临床水平进行验证,同时筛选对相关基因起调控作用的中药。**方法** 应用 GEO 数据库获取 4 个数据集,使用 R 软件包 "Limma"和加权基因共表达网络分析(weighted gene co-expression network analysis,WGCNA)筛选 PCOS 组与健康对照组 的差异表达基因,并对其进行功能富集和细胞免疫浸润分析。应用机器学习算法获取 PCOS 关键基因,绘制列线图,建立受 试者工作特征(receiver operator characteristic, ROC)曲线评估列线图与每个关键基因识别 PCOS 的能力及特异性和敏感性。 收集临床 PCOS 患者外周血单个核细胞,对关键基因表达量与 PCOS 识别情况进行临床验证。通过 COREMINE 数据库、古 今医案云平台预测潜在调控 PCOS 的中药,并分析其性味归经及功效。结果 4 个数据集共获得 42 个样本,其中 PCOS 组 21 个样本,健康对照组 21 个样本。共获得差异基因 127 个,基因本体(gene ontology, GO)分析显示差异基因与肾上腺髓 质素受体信号过程、细胞间桥、类固醇结合等有关;京都基因与基因组百科全书(Kyoto encyclopedia of genes and genomes, KEGG)分析显示差异基因与核因子-κB (nuclear factor-κB, NF-κB)介导的肿瘤坏死因子-α (tumor necrosis factor-α, TNFα) 信号传导、血管生成、白细胞介素-2 (interleukin-2, IL-2) -信号转导和转录激活因子 5 (signal transducer and activator of transcription 5, STAT5)信号等相关。进一步进行细胞免疫浸润,发现 PCOS 组 γδT 细胞、单核细胞、激活的肥大细胞水平 升高,浆细胞、CD4 初始 T 细胞、激活的自然杀伤(natural killer,NK)细胞水平降低。最小绝对收缩和选择算子(logistic least absolute shrinkage and selection operator,LASSO) -Cox 比例风险模型(Cox proportional-hazards model,COX)回归筛选 11 个关键靶点,包括 AK4、DEPP1、DUOX2、FGG、GAREM1、PLOD2、SLC41A2、SPIN4、THNSL1、TMEM187、ZNF443, PCOS 组关键基因的表达量均低于对照组(P<0.05),单个关键基因识别 PCOS 的曲线下面积(area under urve,AUC)为 0.76~0.90, 诺莫图识别 PCOS 的 AUC 为 0.98。临床数据验证共纳入 PCOS 组 12 例, 健康对照组 12 例, PCOS 组 AK4、 ZNF443、DUOX2、DEPP1、FGG、SLC41A2、SPIN4 与 TMEM187 的基因表达量均低于健康对照组 (P<0.05),应用诺莫图 验证对 PCOS 的识别, AUC 为 1。预测到与差异基因相关的中药 85 味,中药的四气以寒、温、平为主,五味以苦、甘、辛 味为主,归经以肝、肺、胃经为主,功效以清热解毒、理气为主。结论 AK4、DEPP1、DUOX2、FGG、GAREM1、PLOD2、 SLC41A2、SPIN4、THNSL1、TMEM187、ZNF443 可能是识别 PCOS 的潜在关键生物标志物以及对潜在治疗中药的预测二 者为 PCOS 的诊断和治疗提供了新思路。

关键词:多囊卵巢综合征;生物信息学、机器学习;临床验证;中药预测;AK4;ZNF443;DUOX2;DEPP1;FGG;SLC41A2;SPIN4;TMEM187

中图分类号: Q811.4; R285 文献标志码: A 文章编号: 0253 - 2670(2025)23 - 8654 - 15

DOI: 10.7501/j.issn.0253-2670.2025.23.017

Exploring key genes of polycystic ovary syndrome based on bioinformatics and machine learning, along with clinical validation and targeted prediction of traditional Chinese medicine

ZHANG Hongyang¹, WANG Qinpeng², WANG Xiaoshan¹, WANG Yongxi³

收稿日期: 2025-08-31

基金项目: 内蒙古自治区自然科学基金项目(2025MS08065); 内蒙古自治区卫生健康科技计划项目(202201553); 赤峰市自然科学基金项目(SZR2025037)

*通信作者: 张红阳 (1992—), 女,回族,博士,主治医师,讲师,研究方向为中医药治疗多囊卵巢综合征等妇科内分泌疾病的研究。
Tel: 15776672045 E-mail: 1508664347@qq.com

- 1. Affiliated Hospital of Chifeng University, Department of Traditional Chinese Medicine, Chifeng 024005, China
- 2. College of Traditional Chinese Medicine (Mongolian Medicine), Chifeng University, Chifeng 024000, China
- 3. College of Traditional Chinese Medicine, Inner Mongolia Medical University, Hohhot 010000, China

Abstract: Objective To explore the key genes of polycystic overy syndrome (PCOS) based on bioinformatics and machine learning methods, and to validate them at the clinical level. At the same time, traditional Chinese medicine that regulates related genes will be screened. Methods A total of four datasets were obtained using GEO database, and Limma and weighted gene co-expression network analysis (WGCNA) analysis methods were used to analyze the differentially expressed genes between PCOS group and healthy control group. Functional enrichment and cellular immune infiltration analysis were performed on them. Apply machine learning algorithms to obtain key genes for PCOS, draw a nomograph, and establish an receiver operator characteristic (ROC) curve to evaluate the ability, specificity, and sensitivity of the nomograph in identifying PCOS with each key gene. Collect peripheral blood mononuclear cells from clinical PCOS patients to clinically validate the expression levels of key genes and PCOS recognition. Predict potential Chinese herbal medicines that regulate PCOS through the COREMININE database and the ancient and modern medical case cloud platform, and analyze their four properties, five flavors, meridian tropism, and efficacy. Results A total of 42 samples are obtained from four datasets, including 21 samples from the PCOS group and 21 samples from the healthy control group. A total of 127 differentially expressed genes are obtained. Gene ontology (GO) analysis showed that the differentially expressed genes are related to adrenomedullin receptor signaling processes, intercellular bridging, steroid binding, etc. Kyoto encyclopedia of genes and genomes (KEGG) analysis showed that the differentially expressed genes are related to nuclear factor- κB (NF- κB) mediated tumor necrosis factor- α (TNF- α) signaling, angiogenesis, interleukin-2 (IL-2)-signal transducer and activator of transcription 5 (STAT5) signaling, etc. Further cellular immune infiltration reveals an increase in levels of gamma delta T cells, monocytes, and activated mast cells in the PCOS group, while levels of plasma cells, CD4 naive T cells, and activated natural killer (NK) cells decreased. Logistic least absolute shrinkage and selection operator (LASSO)-Cox proportional-hazards model (COX) regression screening of 11 key genes, including AK4, DEPP1, DUOX2, FGG, GAREM1, PLOD2, SLC41A2, SPIN4, THNSL1, TMEM187, ZNF443, PCOS the expression levels of key genes in the group are lower than those in the control group (P < 0.05), and the area under urve (AUC) for identifying PCOS by a single key gene is 0.76—0.90, while the AUC for identifying PCOS by nomograph is 0.98. Clinical data validation includs 12 cases in the PCOS group and 12 cases in the healthy control group. The expression levels of AK4, ZNF443, DUOX2, DEPP1, FGG, SLC41A2, SPIN4, and TMEM187 in the PCOS group are lower than those in the healthy control group (P < 0.05). Nomogram validation is used to verify the recognition of PCOS, with an AUC of 1. It is predicted that there will be 85 traditional Chinese medicines related to differential genes. The four qi of traditional Chinese medicine are mainly cold, warmth, and calmness, while the five flavors are mainly bitterness, sweetness, and spicy. The meridian tropism is mainly liver, lung, and stomach, and the efficacy is mainly to heat-clearing, detoxify, and qi-regulating. Conclusion AK4, DEPP1, DUOX2, FGG, GAREM1, PLOD2, SLC41A2, SPIN4, THNSL1, TMEM187, ZNF443 may be potential key biomarker for identifying PCOS and providing new ideas for the diagnosis and treatment of PCOS through the prediction of potential therapeutic herbs.

Key words: polycystic ovary syndrome; bioinformatics; machine learning; clinical validation; traditional Chinese medicine prediction; AK4; ZNF443; DUOX2; DEPP1; FGG; SLC41A2; SPIN4; TMEM187

多囊卵巢综合征(polycystic ovary syndrome, PCOS)为女性常见的生殖内分泌疾病,影响全球8%~18%的女性[1],以月经失调、高雄激素血症、卵巢多囊样改变为核心特征,常伴胰岛素抵抗(IR)、肥胖、心血管代谢风险升高等多系统异常,严重影响女性生殖健康及生命质量。研究表明,PCOS的发生可能与遗传、环境、生活方式等密切相关[2],具体发病机制仍不清。因此,有必要评估有前景的生物标志物,以尽早识别 PCOS 的发生可能。

目前,对于 PCOS 的治疗多以症状为导向,而 多种药物联合使用可能导致不良反应的发生,相比 之下中医药的安全性较高,且具有多成分、多靶点、全身效应的特点,但中医中药治疗以临床辨证为主,尚缺乏统一的治疗方案与处方。因此,本研究综合分析 PCOS 的基因数据集以探索识别 PCOS 的关键基因,同时进行相关中药预测,为 PCOS 的预防和治疗提供新思路。

1 资料与方法

1.1 数据来源

以 "polycystic ovary syndrome"与 "PCOS"为检索词,在基因表达数据库(Gene Expression Omnibus,GEO,https://www.ncbi.nlm.nih.gov/geo/)

中下载 PCOS 基因表达数据集 GSE114419、GSE80432、GSE262735 和 GSE173160。其中GSE114419数据集包括3个PCOS样本与3个健康对照样本,GSE80432数据集包括8个PCOS样本和8个健康对照样本,GSE262735数据集包括4个PCOS样本和4个健康对照样本,GSE173160数据集包括6个PCOS样本和6个健康对照样本。利用R软件包inSilicoMerging对数据集进行合并,进一步使用Evan Johnson等[3]的方法消除批次效应。

1.2 差异基因 (differentially expressed genes, DEGs) 筛选

使用 R 软件包 "Limma" 进行对 "1.1" 项中获得的所有基因进行差异分析,为避免阈值过高遗漏低幅度但有意义的差异基因,以差异倍数>1.2 且 *P*<0.05 为显著差异的筛选标准,将 DEGs 进行热图和火山图的绘制。

1.3 加权基因共表达网络分析(weighted gene co-expression network analysis, WGCNA)

应用 R 软件包"WGCNA"识别 PCOS 中的重 要模块,并建立可视化特征基因网络。利用"1.1" 项中获得所有基因分别计算每个基因的中位数绝 对偏差(median absolute deviation, MAD), 剔除 50% MAD 最小的基因。使用 WGCNA 的 goodSamples Genes 方法去除离群的基因和样本,进一步构建无 标度共表达网络。使用共同表达相似性导出"软" 阈值功率(β)计算邻接,然后将邻接关系转换成拓 扑重叠矩阵(topological overlap matrix, TOM),并 确定基因比率和相异度。根据 TOM 的相异性度量 进行平均连锁等级聚类,基因树状图的最小基因组 大小为30,将具有相似表达谱的基因分类为基因模 块。为了进一步分析模块,计算模块特征基因的相 异性,为模块树状图选择了一条切割线,并合并一 些模块。导入样本临床表型数据,将临床表型与基 因表达矩阵进行匹配,评估各模块的表达模式与 PCOS 表型之间的相关性,并使用热图可视化模块 与表型的关系。

1.4 交集基因分析

将"1.2"项中的 DEGs 与"1.3"项中的模块基 因取交集,绘制韦恩图。

1.5 基因本体 (gene ontology, GO) 功能和京都基因和基因组百科全书 (Kyoto encyclopedia of genes and genomes, KEGG) 通路富集分析

使用 R 软件包 "clusterProfiler"将"1.4"项获

得的交集基因进行 GO 功能与 KEGG 通路富集分析,显著性阈值设置为 P < 0.05,并利用 "ggplot2" 包将分析结果可视化。

1.6 机器学习

为了进一步研究识别 PCOS 的关键基因,使用R 软件包 "glmnet",利用最小绝对收缩和选择算子 (logistic least absolute shrinkage and selection operator,LASSO) -Cox 比例风险模型(Cox proportional-hazards model,COX)方法进行回归分析,设置 3 折交叉验证,以获得最优模型。

1.7 诺莫图构建和受试者工作特征(receiver operator characteristic, ROC) 曲线评估

为了确定关键基因对 PCOS 识别的重要性,使用 R 软件包 "rms"构建诺莫图。诺莫图由"点数"和"总点数"组成,前者表示候选基因的得分,后者表示所有基因得分的总和。通过 ROC 分析进一步评估候选基因和诺谟图的评估价值。ROC 分析曲线下面积(area under urve,AUC)和 95%置信区间(confidence interval,CI),AUC 值>0.7 被认为具有很好的诊断效力。

1.8 免疫浸润分析

使用 R 软件包 "CIBERSORT" 比较 PCOS 组与健康对照组 22 种不同免疫细胞浸润分析,绘制柱状图堆叠用于显示不同样本中每种免疫细胞的比例,箱型图比较两组之间的差异。

1.9 临床试验验证

收集 2022 年就诊于赤峰学院附属医院的 PCOS 患者及同期纳入的健康女性作为健康对照组。于自然月经周期或孕激素撤退出血的第 3~5 天上午空腹采集全血 5 mL,存于 EDTA 抗凝管,3 000 r/min,离心 10 min,弃上清液,取白膜层转移至离心管中冻存,进行高通量测序,测定基因表达量。本研究获得赤峰学院附属医院伦理委员会批准(伦理审批号 fsyy202147),参与者均签署知情同意书。使用 R软件包 "Limma"读取 PCOS 组与健康对照组的表达数据,验证关键基因在 PCOS 组和健康对照组之间的差异,并进行效能验证,绘制 ROC 曲线,计算 AUC 值。

1.10 靶向中药预测分析

在比较毒理基因组学数据库(Comparative Toxicogenomics Database, CTD, http://ctdbase.org/)以支持文献数>10及文献查找 DEGs 对应的化学成分;在 COREMINE 数据库(https://www.coremine.

com/) 中寻找化学成分对应的中药,结合《中国药典》2020 年版与《中药学》2021 版规范化中药名称,在古今医案云平台(https://www.yiankb.com/home)分析中药性能与功效。

2 结果

2.1 多芯片数据集合并

4个数据集共42个样本,其中PCOS组21个样本,健康对照组21个样本。将所有数据集合并归一化,并构建箱线图与均匀流行逼近与投影(uniform manifold approximation and projection,UMAP)图(图1)。箱线图中可以观察到去除批次效应之前各个数据集的样本分布差异较大(图1-A),提示具有批次效应,去除批次效应后各数据集之间的数据分布趋于一致(图1-B),中位数在一条线上。从UMAP图可以观察到去除批次效应之前各个数据集的样本各自聚在一起(图1-C),提示具有批次效应,去除批次效应后各数据集之间的样本互相聚类交织在一起(图1-D),提示较好地去除了批次效应。

2.2 DEGs 鉴定

使用 Limma 方法共鉴定 DEGs 1152 个, 其中 980 个上调、172 个下调。以差异倍数>1.2 且 P<0.05 构

建基因火山图 (图 2-A) 和以显示 30 个基因制作基因 热图 (图 2-B), 其中 400 个上调、72 个下调。

2.3 基因共表达模块鉴定

WGCNA 进行模块加权聚类分析,绘制 PCOS 组和健康对照组的聚类树状图,可见能够完全区分PCOS 组与健康对照组(图 3-A)。为识别与 PCOS 最相关的模块,设置软阈值 β =8,相关性 R^2 >0.86作为标度独立性和平均连通性的标准(图 3-B、C)。设置最小模块大小为 30,敏感性为 3,合并距离小于0.25的模块,最终获得了 5 个共表达模块(图 3-D)。

2.4 模块-特征相关性分析

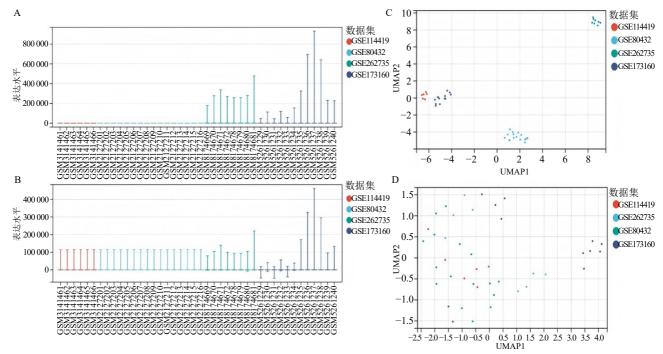
根据软阈值进行 WGCNA,结果显示 darkgreen 模块基因集与 PCOS 最相关,r=-0.56,P<0.001(图 4-A)。特征基因网络热图表明,darkgreen 模块与临床特征高度相关(图 4-B)。在 darkgreen 模块中评估基因显著性,相关系数为 0.61, P<0.001(图 4-C)。

2.5 交集基因鉴定

将 Limma 分析获得的 DEGs 与 WGCNA 中 darkgreen 模块基因取交集, 获得交集基因共 127 个(图 5)。

2.6 GO 功能和 KEGG 通路富集分析

对 127 个交集基因进行 GO 功能和 KEGG 通路



A-去批次前箱型图; B-去批次后箱型图; C-去批次前 UMAP 图; D-去批次后 UMAP 图。

A-box plot before batch removal; B-box plot after batch removal; C-UMAP plot before batch removal; D-UMAP plot after batch removal.

图 1 多数据集整合前后箱型图与 UMAP 图

Fig. 1 Box plots and UMAP plots before and after multi-dataset integration

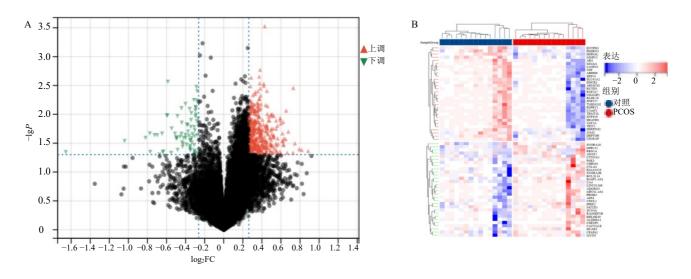
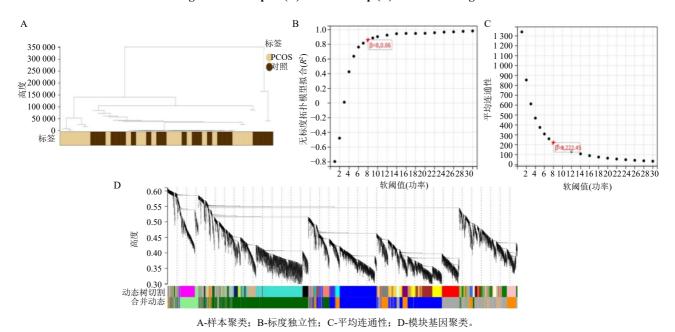


图 2 差异基因火山图 (A) 与热图 (B)

Fig. 2 Volcano plot (A) and heat map (B) of differential genes



A-sample clustering; B-scale independence; C-average connectivity; D-module gene clustering. 图 3 WGCNA 分析的共表达基因模块

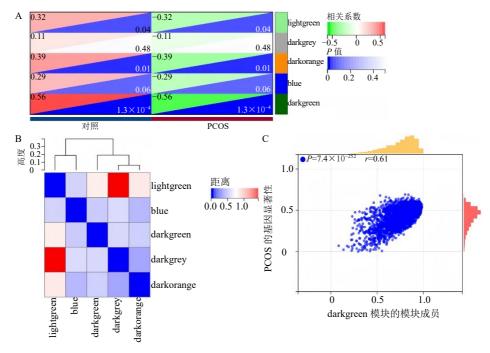
Fig. 3 Co-expressed gene modules analyzed by WGCNA

富集分析。GO 功能富集分析显示,DEGs 生物学过程(biological process,BP)层面与肾上腺髓质素受体信号过程等有关;在细胞组分(cellular components,CC)层面与细胞间桥等有关;在分子功能(molecular function,MF)层面与类固醇结合等有关(图 6-A)。使用 Hallmark 数据库进行 KEGG通路富集分析,结果发现这些 DEGs 主要与核因子-κB(nuclear factor-κB,NF-κB)介导的肿瘤坏死因子-α(tumor necrosis factor-α,TNF-α)信号传导、

血管生成、白细胞介素-2 (interleukin-2, IL-2)-信号转导和转录激活因子 5 (signal transducer and activator of transcription 5, STAT5)信号等相关(图6-B)。

2.7 免疫细胞浸润分析

为探究 PCOS 的免疫微环境,采用 CIBERSORT 反卷积算法比较 22 种免疫细胞在每个样本中的浸润比例(图 7-A)。其中 $\gamma\delta T$ 细胞、单核细胞、激活的肥大细胞在 PCOS 组中高表达(P<0.05、0.01),



A-模块与临床表型相关性热图; B-模块特征向量聚类; C-darkgreen 模块基因显著性。

A-heat map of correlation between module and clinical phenotype; B-module feature vector clustering; C-significance of genes in darkgreen module.

图 4 PCOS 相关模块特征

Fig. 4 Characteristics of PCOS related modules

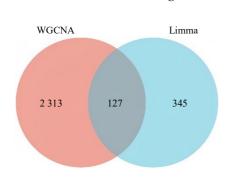


图 5 WGCNA 与 Limma 分析交集基因 Venn 图
Fig. 5 Venn diagram of overlapping genes from WGCNA
and Limma analyses

而浆细胞、CD4 初始 T 细胞、激活的自然杀伤 (natural killer, NK) 细胞在 PCOS 组低表达 (P< 0.05)(图 7-B),提示这些免疫细胞在 PCOS 的发病机制中可能扮演关键角色。

2.8 候选中枢基因鉴定

将 127 个交集基因的表达水平和临床分组用于建立 LASSO 模型,最终共筛选 11 个识别 PCOS 的关键基因,包括腺苷酸激酶 4(adenylate kinase 4,AK4)、蜕膜蛋白前体 1(decidual protein precursors 1,DEPP1)、双氧化酶 2(dual oxidase 2,DUOX2)、纤维蛋白原 γ 链(fibrinogen gamma chain,FGG)、丝裂原活化蛋白激酶 1 相关生长因子受体结合蛋白

2 (activated protein kinase 1 growth factor receptor-bound protein 2, GRB2)调节因子 1(GRB2 associated regulator of MAPK1 1, GAREM1)、前胶原赖氨酸-2-酮戊二酸 5-双加氧酶 2 (procollagen-lysine, 2-oxoglutarate 5-dioxygenase 2, PLOD2)、溶质载体家族 41 成员 2 (solute carrier family 41 member 2, SLC41A2)、纺锤家族成员 4(spindlin family member 4, SPIN4)、苏氨酸合酶样 1 (threonine synthase-like 1, THNSL1)、跨膜蛋白 187 (transmembrane protein 187, TMEM187)、锌指蛋白 443 (zinc finger protein 443, ZNF443) (图 8)。

2.9 诺莫图构建

根据 LASSO 回归分析得出的 11 个关键基因对 PCOS 的危险因素结局贡献大小,构建 PCOS 发生 风险的诺莫图 (图 9)。

2.10 ROC 曲线分析

比较 PCOS 组与健康对照组的 11 个关键基因表达量(图 10-A),建立 ROC 曲线评估诺莫图和每个基因识别 PCOS 的特异性和敏感性,诺莫图的 AUC 为 0.98,单个关键基因识别 PCOS 的 AUC 为 $0.76\sim0.90$,均>0.7,说明所有关键基因对 PCOS 均有较高的识别价值,其中诺模图识别价值最高(图 $10-B\sim M$)。

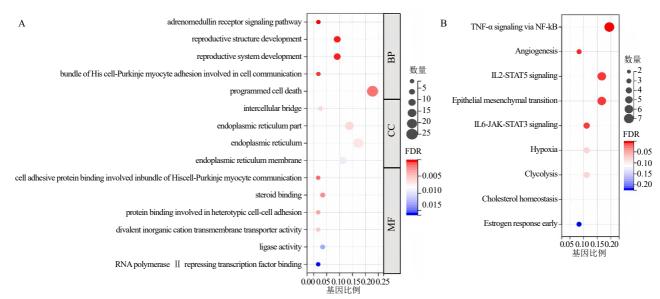
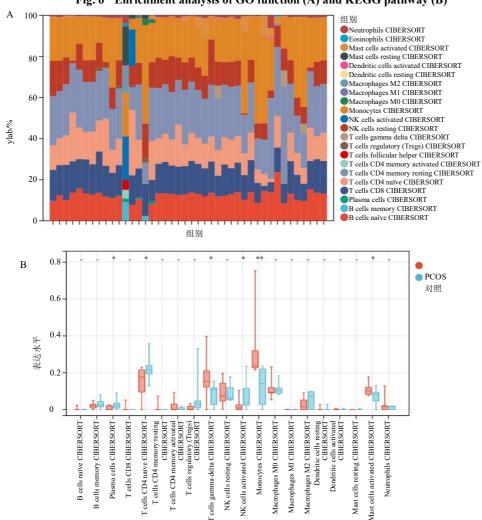


图 6 GO 功能 (A) 和 KEGG 通路 (B) 富集分析

Fig. 6 Enrichment analysis of GO function (A) and KEGG pathway (B)

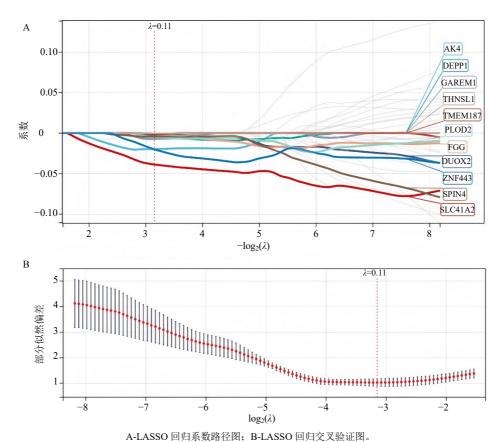


A-不同样品中 22 种免疫细胞比例; B-PCOS 组和健康对照组 22 种免疫细胞比较。

A-proportion of 22 immune cells in different samples; B-comparison of 22 immune cells between PCOS group and healthy control group.

图 7 PCOS 与对照组免疫浸润分析

Fig. 7 Analysis of immune infiltration in PCOS and control groups



A-LASSO regression coefficient path diagram; B-LASSO regression cross validation plot.

图 8 LASSO 模型中的关键基因筛选

Fig. 8 Key gene screening in LASSO model

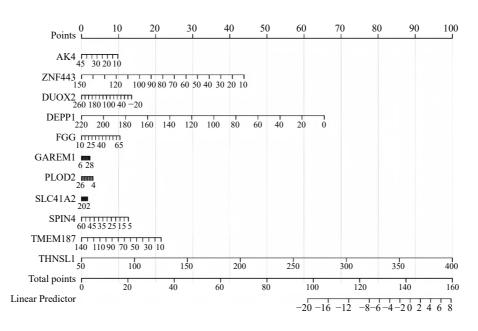


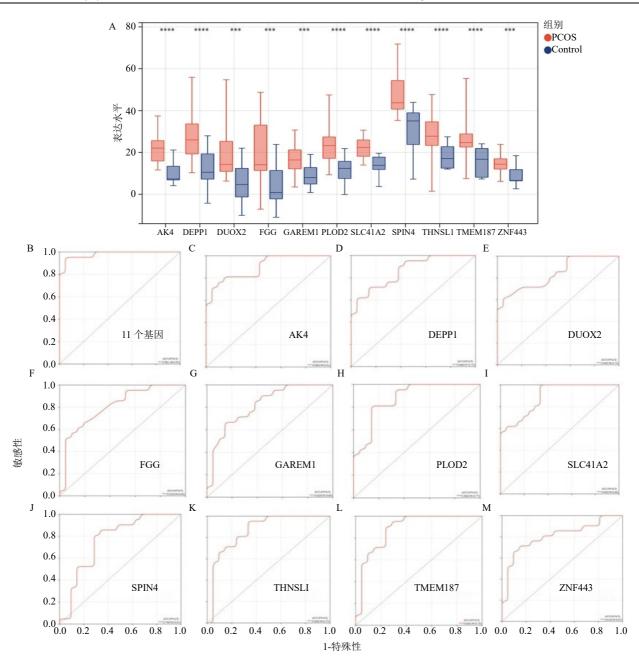
图 9 PCOS 发生的诺莫图识别

Fig. 9 Nomogram recognition of PCOS occurrence

2.11 临床数据验证

回顾性 PCOS 组纳入 12 例,健康对照组 12

例。应用 Limma 对所有基因进行差异表达分析, 共鉴定 DEGs 5 601 个,其中 3 679 个上调、1 922



A-关键基因在 PCOS 组和健康对照组中的表达; B-诺莫图的 ROC 曲线; C~M-每个关键基因的 ROC 曲线。

A-expression of key gene in PCOS group and healthy control group; B-ROC curve of Nomogram; C-M-ROC curve of each key gene.

图 10 关键基因的表达量及 ROC 曲线

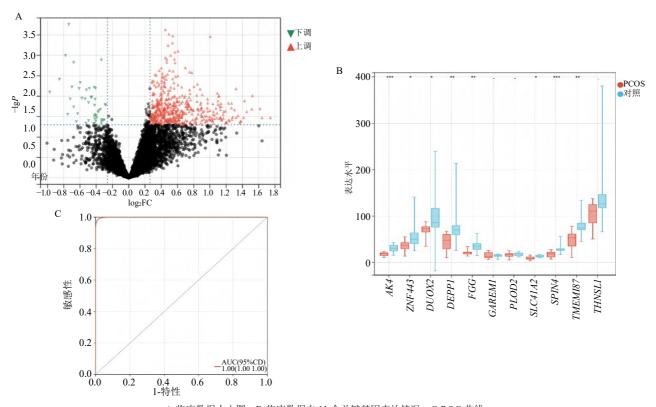
Fig. 10 Expression levels and ROC curves of key genes

个下调。以差异倍数>1.2 构建基因火山图(图 11-A)。比较 PCOS 组与健康对照组的 11 个关键基因表达量(图 11-B),其中 *GAREMI、PLOD2、THNSL1* 3 个基因表达未见差异,可能是训练集数据与验证集数据的样本背景、实验条件、生物学层面的复杂调控性等差异而造成。建立 ROC 曲线评估 11 个关键基因在临床数据中识别 PCOS 的能力,AUC 为 1,提示该 11 个基因联合识别 PCOS

具有较高的价值(图 11-C)。

2.12 靶向中药预测与分析

为能全面匹配中药的复杂作用网络,提升预测的准确性和实用性,本研究将 127 个交集基因提交至 CTD 数据库,通过查阅文献以及检索 CTD 数据库化学成分,共筛选基因 31 个,对应化学成分 104 种;将化学成分提交至 COREMINE 数据库,以 P<0.05 筛选化学成分对应的中药,共筛选



A-临床数据火山图; B-临床数据中 11 个关键基因表达情况; C-ROC 曲线。 A-volcanic plot of clinical data; B-expression status of 11 key genes in clinical data; C-ROC curve.

图 11 临床数据验证

Fig. 11 Clinical data validation

有效化学成分 89 个,对应中药 475 味,出现频次 2 672 次。将频次>10 次以上的中药进行统计,共筛选 83 味中药,高频次中药包括郁金、石斛、人参、淫羊藿、地骨皮等(表 1)。将 475 味中药导入中医药数据挖掘平台进行用药规律分析,显示中药四气以寒、温、平为主,五味以苦、甘、辛味为主,主归肝、肺、胃、脾经(图 12),主要功效以清热解毒、理气、祛风湿、温阳为主(图 13)。

3 讨论

PCOS 临床表现复杂、诊疗周期长、并发症较多,给患者、家庭和社会带来较大的心理与经济负担^[4]。从 1990—2020 年,PCOS 的经济负担估计为每年 80 亿美元^[5]。因此,早筛查、早干预可减少PCOS 的发生发展,有助于保障女性生育健康,并在一定程度上节省医疗成本。

近年来,随着人工智能、多组学技术及中药现代化研究的深入,学者们从不同角度及早识别PCOS的发生^[6]。由于PCOS的诊断标准较多,因此应用临床指标识别PCOS的风险,可能会影响其诊疗。Day等^[7]在大规模全基因组荟萃分析中发现,

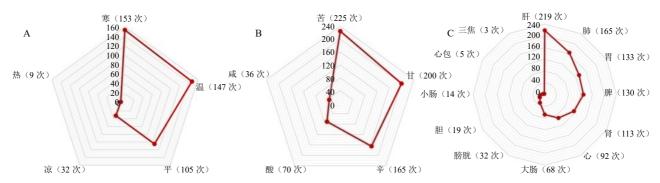
即使不同的 PCOS 诊断标准都具有相同的遗传结构,因此,应用基因预测有助于实现 PCOS 的早期识别,并可为 PCOS 相关药物的研发及中药临床运用提供重要依据,从而提高 PCOS 的临床诊疗水平。

本研究应用生物信息技术分析 4 个公共数据 集, 筛选 PCOS 相关的差异基因。KEGG 分析表明, PCOS 相关差异基因主要富集在 NF-κB 介导的 TNF-α 信号传导、血管生成、IL-2-STAT5 信号通路 等。可见其多与炎症相关,进一步进行免疫浸润分 析,发现 PCOS 组 γδT 细胞、单核细胞、激活的肥 大细胞水平升高,浆细胞、CD4 初始 T 细胞、激活 的 NK 细胞水平降低。γδT 细胞是一种独特而罕见 的 T 细胞群体, 在皮肤、脂肪组织等外周组织中聚 集。γδT 细胞在胸腺发育过程中发育成效应细胞, 这种早期效应程序促使其对炎症反应极为灵敏, 促 进γ干扰素 (interferon-γ, INF-γ)、IL-17A 和 IL-23 的分泌[8-9]。研究发现, γδT 细胞可能影响 PCOS 的 发生[10]。单核细胞可分泌炎性细胞因子介导炎症反 应的发生。研究发现,单核细胞分泌 TNF-α、IL-6, 诱导 PCOS 胰岛素抵抗 (insulin resistance, IR), 刺

表 1 高频中药

Tabla 1	High frequency	traditional	Chinasa ma	dicina
IADICI	III2II II CUUCIICY	u auiuviiai	Chinese me	uitiit

中药	频数	中药	频数	中药	频数
郁金	30	姜黄	16	莱菔子	13
石斛	29	灵芝	16	芦根	13
人参	28	西洋参	16	三七叶	13
人参叶	28	大蒜	16	桑叶	13
淫羊藿	24	丹参	15	茵陈	13
地骨皮	23	茯神	15	板蓝根	12
甘草	21	红参	15	苍术	12
玉米须	21	黄芪	15	诃子	12
白果	20	黄芩	15	厚朴花	12
大黄	20	山楂	15	槐花	12
蜂蜜	20	赤芍	14	火麻仁	12
干姜	20	何首乌	14	麻黄	12
黑芝麻	20	连翘	14	女贞子	12
木香	20	肉桂	14	桑白皮	12
三七	20	五味子	14	莪术	11
麝香	20	紫苏	14	葛根	11
生姜	20	车前子	13	厚朴	11
瓦楞子	20	大青叶	13	黄精	11
地黄	19	当归	13	荔枝核	11
旋覆花	19	海藻	13	龙胆	11
浮小麦	18	红花	13	马齿苋	11
枸杞子	18	胡荽	13	牡蛎	11
浮萍	17	芦根	13	桑枝	11
菟丝子	17	三七叶	13	石决明	11
白芥子	16	桑叶	13	石榴皮	11
茯苓	16	茵陈	13	桃仁	11
黑种草子	16	板蓝根	12	紫苏梗	11
胡芦巴	16	蒺藜	13		



A-四气雷达图; B-五味雷达图; C-归经雷达图

A- radar diagram of four properties; B-radar diagram of five flavor; C-radar diagram of meridian tropism.

图 12 四气、五味、归经雷达图

Fig. 12 Radar diagrams of four properties, five flavors, and meridian tropism

激雄激素的产生,并引起下丘脑-垂体-卵巢轴分泌紊乱。肥大细胞是一种在全身广泛分布的固有免疫细胞,肥大细胞可以产生细胞因子,如 TNF-α、IL-31、IL-6等,促进临床炎症的发生。此外,其脱颗粒释放的介质如类胰蛋白酶、糜蛋白酶、半胱氨酰白三烯等均可诱导炎症的发生[11]。可见 γδT 细

胞、单核细胞、肥大细胞的增加可能是 PCOS 慢性 炎症过程的起始因素。NK 细胞是天然免疫系统的核 心组成部分,能够快速识别肿瘤、病毒感染。目前关 于 NK 细胞与 PCOS 的关系研究尚不明确,Li 等[12] 研究发现,NK 细胞在 PCOS 中所占比例较低,该结 果与本研究一致。然而,也有研究发现,PCOS 中 NK

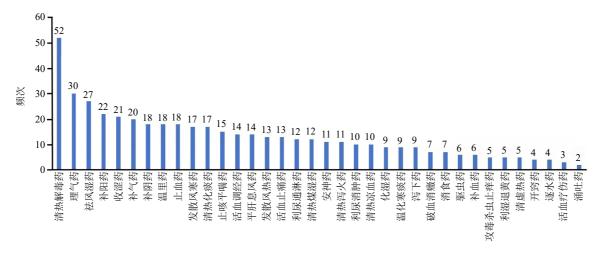


图 13 药物分类柱形图

Fig. 13 Bar chart of drug classification

细胞水平高于对照组或两组之间无差异^[13-14],因此仍需要深入探索。CD4 初始 T 细胞及浆细胞等免疫细胞在 PCOS 中同样扮演着重要角色,但与其之间的关系仍需进一步研究。

经机器学习构建诺莫图及 ROC 曲线分析验证, 确定 PCOS 相关的关键基因包括 AK4、DEPP1、 DUOX2, FGG, GAREM1, PLOD2, SLC41A2, SPIN4、THNSL1、TMEM187、ZNF443,将每个基 因的表达量转化为分数,分数关联线性预测。当线 性预测值较高时,可以对 PCOS 进行早期监测和干 预。AK4 属于腺苷酸激酶家族,其在细胞能量代谢 和维持各种亚细胞的腺嘌呤核苷酸稳态中起着关 键作用。在 PCOS 中尚缺乏研究, 但在子宫内膜癌 中 AK4 表达显著较高[15]。DEPP1 是由孕酮、雄激 素和缺氧诱导的在卵巢、脂肪组织和肝脏中高表达 的自噬调节因子[16],自噬与 PCOS 排卵障碍、高雄 激素、IR、肥胖、内分泌紊乱等密切相关[17]。DUOX2 基因的突变是导致甲状腺功能减退的主要原因,而 甲状腺功能减退是导致 PCOS 的原因之一[18-19]。 FGG 基因的突变会导致低纤维蛋白原血症,从而影 响排卵后卵巢组织修复[20]。另外, FGG 表达量降低 影响早期胚胎的附植[21]。GAREM 是一种与多种细 胞生长因子受体信号相关的衔接蛋白,参与细胞增 殖和肿瘤发生。研究发现, GAREM1 上调丝裂原活 化蛋白激酶 (mitogen-activated protein kinase, MAPK) 途径与磷脂酰肌醇-3-羟激酶 (phosphatidylinositol-3-hydroxykinase, PI3K)途径^[22]。 PLOD2 在肺癌、胃癌、卵巢癌组织中高表达,且可 通过激活 PI3K/蛋白激酶 B(protein kinase B,Akt)/ 哺乳动物雷帕霉素靶蛋白(mammalian target of rapamycin, mTOR)途径促进糖酵解^[23-24]。SLC41A2 的遗传变异显著影响血清镁介导的糖尿病风险^[25],此外,SLC41A2 的表达增加了瞬时受体电位黑素瘤相关蛋白 7(transient receptor potential melastatin 7,TRPM7)敲除细胞对凋亡的敏感性,并促进细胞的氧化应激^[26]。SPIN4 结合特定的组蛋白修饰促进Wnt 信号传导^[27]。TMEM187 过表达降低细胞清除活性氧,进而引起细胞氧化应激。然而,过表达的TMEM187 也可促进细胞凋亡,抑制肿瘤的形成与生长^[28]。目前对 THNSL1 与 ZNF443 基因研究较少,研究发现,THNSL1 在弥漫大 B 细胞淋巴瘤的预测与治疗中具有一定的价值^[29]。ZNF443 与鼻咽癌患者的生存、艾滋病免疫等相关^[30-31]。

PCOS 在中医学中属于"月经后期""闭经""不孕症"等范畴,为本虚标实之证,肾虚为本,脾虚肝郁由本及标,痰湿血瘀为标^[32]。中医认为肾为先天之本,主生殖,肾气的发育与女性的生殖密切相关。肾气不足,冲任失调,血海不能按时满溢,出现月经不调。肾阴不足,卵泡失于滋养;肾阳不足,卵泡无力排出^[33],故肾虚为 PCOS 月经失调、不孕的根本原因。脾为后天之本,气血生化之源。脾失健运,则水湿内停,湿聚成痰,阻滞气机,困遏脾阳。肝主疏泄,调畅全身气机,肝失疏泄,影响冲任气血条畅。痰湿血瘀为 PCOS 的重要病理产物,痰湿阻滞冲任、胞宫,导致月经失调、不孕等;泛溢肌肤,导致肥胖、痤疮等;郁久化热,形成湿热,导致炎症和代谢紊乱^[34]。瘀血阻滞于胞宫胞络,气血运行不畅,导致卵巢包膜增厚,卵子难以排出,

久而形成窠囊[35]。以上因素相互联系、相互影响, 形成了 PCOS"本虚标实,虚实夹杂"的复杂病理 网络。本研究总结预测治疗 PCOS 的潜在中药,其 功效主要以清热解毒、理气、祛风湿、温阳为主, 切中 PCOS 以肾虚、脾虚、肝郁导致气血失调,气 郁湿阻日久, 化瘀化火, 湿热内生的病理特点。治 疗上,通过温补肾阳、滋补肾阴、健脾化湿、疏肝 理气、清热除湿,达到改善症状、缓解炎症的目的, 此为中医药治疗 PCOS 的关键。从药性来看,四气 以寒、温、平为主, 五味以苦、甘、辛味为主, 归 经为肝、肺、胃、脾经,与 PCOS 的病机相吻合。 苦寒药多能清热燥湿,清上焦湿热、肝经湿热、肠 道湿热[36-38],改善临床症状、抑制炎症水平、调节 肠道菌群。甘温多能健脾益气、增强脾胃功能,达 到祛湿化痰的目的;同时,甘温益气以祛除郁热, 为非肥胖型 PCOS 的主要治疗方法[39]。《素问·至真 要大论篇》云:"阳明之复,治以辛温,佐以苦甘, 以苦泄之,以苦下之"。辛味药则具有发散行气的作 用,与苦味药同用,一辛一苦,一开一降,调畅中焦 之气。因此,"辛开苦降"也为 PCOS 的重要治法[40]。 郁金、石斛、人参、淫羊藿、地骨皮等为预测治疗 PCOS 的高频次中药。郁金可以降低 TNF-α 水平, 通过调控 PI3K/Akt/mTOR 途径改善 PCOS 患者的 IR[41]。石斛能够升高 PCOS 患者促卵泡生成素 (follicle-stimulating hormone, FSH) 和孕酮 (progesterone, P)水平,降低黄体生成素(luteinizing hormone, LH)与 FSH 比值,同时通过激活 PI3K/Akt 通路,增加颗粒细胞的凋亡[42]。Hu等[43]研究发现, 石斛多糖通过激活沉默信息调节因子 2 (sirtuin 2, SIRT2)调节 PCOS 颗粒细胞的 IR,恢复糖酵解途 径。人参皂苷可激活棕色脂肪组织预防 PCOS 体质 量增加,此外,人参皂苷可以改善 PCOS 高雄激素 血症,减少囊性卵泡的数量[44]。对 PCOS 用药规律 分析发现,淫羊藿-巴戟天为高频聚类方[45]。淫羊藿 有效成分淫羊藿苷通过抑制 PCOS 大鼠卵巢颗粒细 胞凋亡[46],促进卵泡发育、成熟、排卵[47],以及改 善PCOS肝脂肪变性等[48]。地骨皮煎煮液调控PI3K/ 蛋白激酶 B(protein kinase B, PKB) 途径改善 PCOS 大鼠动情周期、降低血清激素水平与体质量[49]。

综上所述,AK4、DEPP1、DUOX2、FGG、GAREM1、PLOD2、SLC41A2、SPIN4、THNSL1、TMEM187、ZNF443 是识别 PCOS 的关键基因。该基因既保证 PCOS 识别的准确性,又能保证临床的

可取性,可为 PCOS 的早期识别提供更为便捷、经济的手段。在治疗方面,免疫细胞可作为 PCOS 的作用靶点,具有较高的创新性与应用前景。此外,对 PCOS 治疗中药的预测有助于临床精准化治疗及为新药物的研发提供依据。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] Zhang J C, Zhang H L, Xin X Y, *et al.* Efficacy of flavonoids on animal models of polycystic ovary syndrome: A systematic review and meta-analysis [J]. *Nutrients*, 2022, 14(19): 4128.
- [2] 葛亚杰,徐文,关诗敏,等. 多囊卵巢综合征病因及其 发病机制的研究进展 [J]. 吉林大学学报: 医学版, 2024,50(1): 288-294.
- [3] Evan Johnson W, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods [J]. *Biostatistics*, 2007, 8(1): 118-127.
- [4] 王佳卉,马增香.多囊卵巢综合征及其并发症的直接 经济负担研究进展 [J]. 卫生经济研究,2022,39(3):24-27.
- [5] Zhang J C, Zhu Y T, Wang J H, et al. Global burden and epidemiological prediction of polycystic ovary syndrome from 1990 to 2019: A systematic analysis from the Global Burden of Disease Study 2019 [J]. PLoS One, 2024, 19(7): e0306991.
- [6] Guan G X, Pu Y F, Gao Y, *et al.* Progress of the application clinical prediction model in polycystic ovary syndrome [J]. *J Ovarian Res*, 2023, 16(1): 230.
- [7] Day F, Karaderi T, Jones M R, et al. Large-scale genomewide meta-analysis of polycystic ovary syndrome suggests shared genetic architecture for different diagnosis criteria [J]. PLoS Genet, 2018, 14(12): e1007813.
- [8] Chen D, Guo Y L, Jiang J H, et al. γδ T cell exhaustion: Opportunities for intervention [J]. J Leukoc Biol, 2022, 112(6): 1669-1676.
- [9] Sun L N, Su Y H, Jiao A J, et al. T cells in health and disease [J]. Sig Transduct Target Ther, 2023, 8: 235.
- [10] Gao M G, Liu X H, Du M X, et al. Identification of immune cell infiltration and effective biomarkers of polycystic ovary syndrome by bioinformatics analysis [J]. BMC Pregnancy Childbirth, 2023, 23(1): 377.
- [11] Parente R, Giudice V, Cardamone C, et al. Secretory and membrane-associated biomarkers of mast cell activation and proliferation [J]. Int J Mol Sci, 2023, 24(8): 7071.
- [12] Li S M, Li Y M, Sun Y, et al. Deconvolution at the singlecell level reveals ovarian cell-type-specific transcriptomic changes in PCOS [J]. Reprod Biol Endocrinol, 2024,

22(1): 24.

- [13] Yuan Y, Mao Y, Yang L, *et al.* Analysis of macrophage polarization and regulation characteristics in ovarian tissues of polycystic ovary syndrome [J]. *Front Med*, 2024, 11: 1417983.
- [14] Torstensson S, Ascani A, Risal S, et al. Androgens modulate the immune profile in a mouse model of polycystic ovary syndrome [J]. Adv Sci, 2024, 11(28): e2401772.
- [15] Lu W, Yang Z Y, Wang M J, *et al.* Identification and verification of AK4 as a protective immune-related biomarker in adipose-derived stem cells and breast cancer [J]. *Heliyon*, 2024, 10(7): e27357.
- [16] Xu C, Liao M X, Zhang S Z, et al. The comorbidity of depression and diabetes is involved in the decidual protein induced by progesterone 1 (Depp1) dysfunction in the medial prefrontal cortex [J]. Metabolites, 2025, 15(1): 34.
- [17] Kumariya S, Ubba V, Jha R K, *et al.* Autophagy in ovary and polycystic ovary syndrome: Role, dispute and future perspective [J]. *Autophagy*, 2021, 17(10): 2706-2733.
- [18] Hesampour A, Jafarabadi M, Rajabi S, et al. Dual oxidase 1, 2 gene expression in women with polycystic ovary syndrome (PCOS) [J]. J Family Reprod Health, 2023, 17(4): 205-215.
- [19] Singla R, Gupta Y, Khemani M, et al. Thyroid disorders and polycystic ovary syndrome: An emerging relationship [J]. Indian J Endocrinol Metab, 2015, 19(1): 25-29.
- [20] 蒋旭东, 刘宇, 邬建飞, 等. 牦牛FGG组织表达与雌性 生殖器官中定位分析 [J]. 生物技术通报, 2022, 38(11): 286-294.
- [21] Lecce L, Kaneko Y, Madawala R J, et al. ICAM1 and fibrinogen-γ are increased in uterine epithelial cells at the time of implantation in rats [J]. Mol Reprod Dev, 2011, 78(5): 318-327.
- [22] Voynova E, Kulebyakin K, Grigorieva O, et al. Declined adipogenic potential of senescent MSCs due to shift in insulin signaling and altered exosome cargo [J]. Front Cell Dev Biol, 2022, 10: 1050489.
- [23] Ye W C, Zhang X L, Tang Z J, *et al*. Comprehensive analysis of glycometabolism-related genes reveals PLOD2 as a prognostic biomarker and therapeutic target in gastric cancer [J]. *BMC Gastroenterol*, 2025, 25(1): 256.
- [24] 郑朗,陈春英,沈栗帆,等. 卵巢癌组织中 PLOD2 的 表达及其在缺氧条件下通过 PI3K/Akt 通路对卵巢癌细 胞作用的研究 [J]. 中国性科学, 2022, 31(2): 75-79.
- [25] Kieboom B C T, Ligthart S, Dehghan A, *et al.* Serum magnesium and the risk of prediabetes: A population-based cohort study [J]. *Diabetologia*, 2017, 60(5): 843-

853.

- [26] Chen H C, Su L T, González-Pagán O, *et al.* A key role for Mg²⁺ in TRPM7's control of ROS levels during cell stress [J]. *Biochem J*, 2012, 445(3): 441-448.
- [27] Lui J C, Wagner J, Zhou E, *et al.* Loss-of-function variant in SPIN4 causes an X-linked overgrowth syndrome [J]. *JCI Insight*, 2023, 8(9): e167074.
- [28] 郝双影. TMEM187 在线粒体铁代谢中的作用 [D]. 南京: 南京大学, 2016.
- [29] Zhou Z Z, Lu J C, Guo S B, *et al.* A mitochondria-related signature in diffuse large B-cell lymphoma: Prognosis, immune and therapeutic features [J]. *Cancer Med*, 2025, 14(2): e70602.
- [30] Liu G H, Zeng X J, Wu B L, et al. RNA-Seq analysis of peripheral blood mononuclear cells reveals unique transcriptional signatures associated with radiotherapy response of nasopharyngeal carcinoma and prognosis of head and neck cancer [J]. Cancer Biol Ther, 2020, 21(2): 139-146.
- [31] 彭小青. 数据挖掘艾滋病免疫相关关键基因及潜在药物预测 [D]. 南宁: 广西中医药大学, 2022.
- [32] 郁悦, 陈彤, 征征, 等. 口服中成药治疗多囊卵巢综合 征有效性与安全性的网状 Meta 分析 [J]. 中草药, 2022, 53(23): 7477-7490.
- [33] 刘颖, 闫颖, 张晗. 哈孝廉教授治疗多囊卵巢综合征排卵障碍性不孕临证经验 [J/OL]. 辽宁中医杂志, (2025-08-06) [2025-08-27]. https://link.cnki.net/urlid/21.1128.R. 20250806. 0922.010.
- [34] 李凤英, 杨华, 孙哲, 等. 王琪教授诊治多囊卵巢综合征学术经验探析 [J/OL]. 中华中医药学刊, (2025-05-22) [2025-08-27]. https://link.cnki.net/urlid/21.1546.R. 20250522.1308.014.
- [35] 逯克娜, 黄金华, 林寒梅. 当代中医妇科名家辨治多囊 卵巢综合征的学术思想荟萃 [J]. 时珍国医国药, 2024, 35(8): 1994-1997.
- [36] 张金枫, 胡荣魁, 任苏妍, 等. 基于三焦气化应用苦辛 开泄法调治湿热型多囊卵巢综合征 [J]. 环球中医药, 2023, 16(4): 727-730.
- [37] 肖淑兰, 杜彦懿, 王嘉玥. 龙胆泻肝汤加减治疗肝经湿热型 PCOS 的疗效及对血清性激素、炎症因子的影响 [J]. 中药材, 2024, 47(1): 226-230.
- [38] 喻星星. 湿热型多囊卵巢综合征患者肠道菌群分析 [D]. 南昌: 江西中医药大学, 2021.
- [39] 康瀚卿, 黄灿灿, 岳斌, 等. 基于"火与元气不两立" 论非肥胖型多囊卵巢综合征 [J]. 上海中医药杂志, 2025, 59(4): 30-34.
- [40] 罗献英,杨清,谢秀超.辛开苦降法对多囊卵巢综合征 胰岛素抵抗患者卵巢功能、子宫内膜容受性及磷脂酰

- 肌醇 3 激酶/丝苏氨酸蛋白激酶通路的影响 [J]. 世界中西医结合杂志, 2024, 19(5): 987-992.
- [41] Yu H, Zhang L, Yin F, et al. Exploring the molecular targets and mechanism of *S. miltiorrhiza-C. aromatica* in treating polycystic ovary syndrome based on network pharmacology [J]. *Ann Transl Med*, 2023, 11(3): 149.
- [42] Gao J F, Huang Y, Song Y Y, *et al.* Therapeutic effects of *Althaea officinalis* L. and metformin on estradiol-induced polycystic ovary syndrome in rats: Insights into the PI3K/AKT pathway, inflammation, and oxidative stress [J]. *Cell J*, 2024, 26(8): 473-486.
- [43] Hu R, Nong W H, Huo P, et al. Dendrobium nobile-derived polysaccharides stimulate the glycolytic pathway by activating SIRT2 to regulate insulin resistance in polycystic ovary syndrome granulosa cells [J]. Int J Biol Macromol, 2024, 278(Pt 2): 134780.
- [44] Ye R C, Yan C L, Zhou H Q, et al. Brown adipose tissue

- activation with ginsenoside compound K ameliorates polycystic ovary syndrome [J]. *Br J Pharmacol*, 2022, 179(18): 4563-4574.
- [45] 高芷铭. 基于隐结构模型的多囊卵巢综合征中医证候分析及其用药规律研究 [D]. 长春: 长春中医药大学, 2023.
- [46] 徐琼芳, 钟斐, 李子帅. 淫羊藿苷调节 SDF-1/CXCR4 信号通路对多囊卵巢综合征大鼠卵巢颗粒细胞凋亡的 影响 [J]. 天津医药, 2024, 52(7): 727-732.
- [47] 王小洁, 张云锋, 张彦敏. 淫羊藿苷基于 GDF-9/Smads 对多囊卵巢小鼠排卵的作用 [J]. 西北药学杂志, 2023, 38(5): 65-70.
- [48] 海洋. 淫羊藿苷改善多囊卵巢综合征合并非酒精性脂肪肝机制研究 [D]. 重庆: 重庆医科大学, 2023.
- [49] 张婷, 郭建友. 地骨皮煎煮液对多囊卵巢大鼠胰岛素信号传导途径中 PI3K/PKB 分子表达的影响 [J]. 中国中药杂志, 2015, 40(10): 2004-2008.

[责任编辑 潘明佳]