

DOI: 10.3969/j.issn.2096-6113.2024.01.007

引用格式:张慧杰,曲洪美,马江磊.基于加权基因共表达网络分析鉴定 PE 患者的潜在生物标志物[J].巴楚医学,2024,7(1):44-49.

# 基于加权基因共表达网络分析鉴定 PE 患者的潜在生物标志物

张慧杰<sup>1</sup> 曲洪美<sup>1</sup> 马江磊<sup>2</sup>

(1. 烟台毓璜顶医院 产科, 山东 烟台 264000; 2. 烟台市奇山医院 感染性疾病科, 山东 烟台 264000)

**摘要:** **目的:**本研究旨在通过对子痫前期(PE)患者基因数据集进行加权基因共表达网络分析(WGCNA),寻找 PE 的分子标志物。**方法:**从 GEO 数据库下载 GSE10588 表达数据集,该数据集包括 17 例 PE 患者和 26 例对照者的胎盘组织转录组数据。使用差异表达分析及 WGCNA 筛选与 PE 相关的重点差异表达基因(DEGs),利用基因本体论(GO)和京都基因与基因组百科全书(KEGG)通路分析来探索重点基因的生物学作用。构建蛋白-蛋白相互作用(PPI)网络筛选关键枢纽基因,并通过风险预测模型和受试者工作特征(ROC)曲线评价这些枢纽基因对 PE 的预测效果。**结果:**共获得 138 个 DEGs,其中包括 100 个上调和 38 个下调 DEGs。通过 WGCNA 确定了 23 个与 PE 发病最相关的重点 DEGs。GO 和 KEGG 通路分析显示,这些重点 DEGs 在调节促性腺激素分泌、白细胞分化、造血功能、内分泌功能及 B 细胞的激活等生物学作用中发挥重要作用。在 PPI 网络中,卵泡抑素相关基因 3(FSTL3)、酪氨酸激酶受体 1(FLT1)、抑制素亚单位 A(INHBA)、N-myc 下游调节基因 1(NDRG1)和瘦素(LEP)等成为节点最多的基因。风险预测和 ROC 曲线表明,FSTL3 和 LEP 具有较高的风险得分和诊断价值。**结论:**FSTL3 和 LEP 可以作为预测和诊断 PE 的生物标志物。

**关键词:** 子痫前期; 加权基因共表达网络分析; 生物标志物; 卵泡抑素相关基因 3; 瘦素  
**中图分类号:** R714.25 **文献标志码:** A **文章编号:** 2096-6113(2024)01-0044-06

开放科学(资源服务)标识码(OSID):



## Identification of Potential Biomarkers in Patients with Preeclampsia Based on Weighted Gene Co-Expression Network

Zhang Huijie<sup>1</sup> Qu Hongmei<sup>1</sup> Ma Jianglei<sup>2</sup>

(1. Department of Obstetrics, Yantai Yuhuangding Hospital, Yantai 264000, China; 2. Department of Infectious Diseases, Yantai Qishan Hospital, Yantai 264000, China)

**Abstract Objective:** To search for molecular markers of preeclampsia (PE) by using weighted gene co-expression network analysis (WGCNA) on a genetic dataset of patients with preeclampsia (PE). **Methods:** The GSE10588 expression dataset was downloaded from the GEO database and included placental tissue transcriptome data from 17 PE patients and 26 controls. Differential expression analysis and WGCNA were used to screen key differentially expressed genes (DEGs) associated with PE, and gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis were used to explore the biological roles of key genes. The protein-protein interaction (PPI) network was constructed to screen key hub genes, and the predictive effect of these hub genes on PE was evaluated by risk prediction model and receiver

基金项目:山东省自然科学基金项目(No: ZR2021MH223)

作者简介:张慧杰,女,硕士,主治医师,研究方向为产前诊断。E-mail: yhdzhj@126.com

通信作者:马江磊,男,硕士,主治医师,研究方向为分子诊断。E-mail: mjl231208@126.com

operating characteristic (ROC) curve. **Results:** A total of 138 DEGs were obtained, including 100 up-regulated and 38 down-regulated DEGs. The WGCNA identified 23 key DEGs that were most associated with the onset of PE. GO and KEGG pathway analysis showed that these key DEGs play an important role in regulating gonadotropin secretion, leukocyte differentiation, hematopoietic function, endocrine regulation, and B-cell activation. In the PPI network, follistatin-related gene 3 (FSTL3), tyrosine kinase receptor 1 (FLT1), statin subunit A (INHBA), N-myc downstream regulatory gene 1 (NDRG1) and leptin (LEP) were the genes with the most nodes. Risk prediction and ROC curve showed that FSTL3 and LEP had high risk scores and diagnostic value. **Conclusion:** FSTL3 and LEP can be used as biomarkers for the prediction and diagnosis of PE.

**Keywords** preeclampsia; weighted gene coexpression network analysis(WGCNA); biomarkers; follistatin-related gene 3(FSTL3); leptin(LEP)

子痫前期(preeclampsia, PE)是一种妊娠期的特异性疾病,约占所有妊娠疾病的3%~5%<sup>[1]</sup>。该疾病的特征是高血压引起血管内皮功能障碍,导致肝脏、肾脏、大脑以及胎盘等多器官损伤。在全世界范围内,PE是孕产妇疾病发生和死亡的重要原因之一<sup>[2]</sup>。此外,PE还容易引发早产,导致新生儿疾病的发生和围产期死亡等问题。因此,PE生物标志物的筛选有助于疾病的早期诊断和及时治疗,对于提高孕产妇及新生儿的存活率具有重要意义。

可溶性纤维样酪氨酸激酶1(soluble fms-like tyrosine kinase 1, sFlt-1)和胎盘生长因子(placental growth factor, PlGF)的检测已在临床广泛应用。孕周<37周时,若sFlt-1:PlGF≤38,可排除后一周内发生PE的可能性,该指标阴性预测值可达99.3%,敏感性为80.0%,特异性为78.3%;若sFlt-1:PlGF>38,则其预测4周内发生PE的阳性准确值只有36.7%,敏感性为66.2%,特异性为83.1%,不足以满足临床需要<sup>[3-4]</sup>。

为了寻找更有效的PE预测和诊断生物标志物,本研究利用基因表达综合(gene expression omnibus, GEO)数据库的PE患者和健康对照者的基因表达谱进行分析,通过差异表达基因(differentially expressed genes, DEGs)筛选和加权基因共表达网络分析(weighted gene co-expression network analysis, WGCNA),获得疾病相关重要基因。WGCNA在转录组学、蛋白质组学、遗传组学和代谢组学等领域都得到广泛应用,并可用于筛选候选生物标志物或治疗靶点<sup>[5-6]</sup>。本研究旨在筛选PE生物标志物,以期作为PE患者的早期诊断及治疗提供参考依据。

## 1 资料与方法

### 1.1 资料来源

从GEO数据库(<https://www.ncbi.nlm.nih.gov/geo/>)中的搜索界面输入“preeclampsia”,下载数据集GSE10588及其相应的平台文件GPL2986。该数据集中包含17例PE患者和26例对照者的胎盘组织转录组测序数据。

### 1.2 DEGs 鉴定

采用R 4.2.2软件中的“limma”数据包进行DEGs的分析和筛选<sup>[7]</sup>。筛选标准为:差异倍数(fold change, FC)的绝对值取对数≥1及矫正后的 $P \leq 0.05$ 。使用“pheatmap”和“ggplot2”R包进行DEGs的可视化,生成火山图和热图。

### 1.3 构建 WGCNA 网络

使用“WGCNA”R包构建PE患者WGCNA网络<sup>[5]</sup>,并评估不同模块与PE发病机制的相关性,选择与PE疾病最相关的模块作为WGCNA基因集团做后续分析。

### 1.4 重点基因的筛选

使用“VennDiagram”R包获得DEGs集团与WGCNA产生的与疾病最相关基因集团的交集基因,这些交集基因被认为是与PE发病机制相关的关键DEGs。

### 1.5 功能富集分析

使用“clusterProfiler”R包进行基因本体论(gene ontology, GO)和京都基因与基因组百科全书(kyoto encyclopedia of genes and genomes, KEGG)通路富集分析,以了解潜在重点基因在PE发生发展中的生物学机制<sup>[8]</sup>。

### 1.6 构建 PPI 网络和枢纽基因的鉴定

使用STRING在线工具(<https://cn.string-db.org/>)预测和可视化分子的相互作用和PPI网络<sup>[9]</sup>。将重点基因输入到STRING工具中,得到PPI网络。之后,将获得的PPI网络导入Cytoscape软件(版本号:3.9.1)中,并使用Degree算法计算出节点最多的前10个基因,选择前5个基因作为关键枢纽基因进行评估。

### 1.7 构建风险预测模型及诊断评估

风险预测列线图模型是通过计算哈雷尔一致性指数来评估对危险因素的预测风险<sup>[10]</sup>。使用“rms”R 包进行枢纽基因与 PE 的风险预测<sup>[11]</sup>。使用“pROC”R 包构建受试者工作特征(receiver operator characteristic, ROC)曲线以评估枢纽基因的诊断有效性<sup>[12]</sup>。ROC 曲线下面积(area under the curve, AUC)用于评估准确度,一般定义为  $0.9 \leq AUC < 1$  表示良好的准确性。

## 2 结果

### 2.1 DEGs 鉴定

根据 DEGs 筛选标准,对数据集 GSE10588 分析后,总共获得 138 个 PE 患者与对照者的 DEGs,包括个 100 上调和 38 个下调的 DEGs,见图 1。

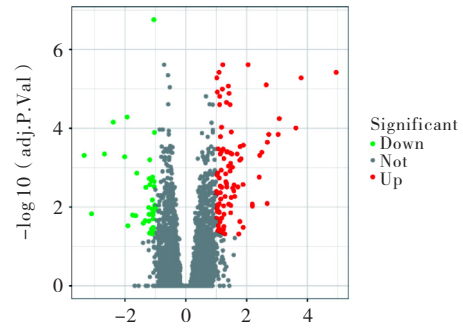
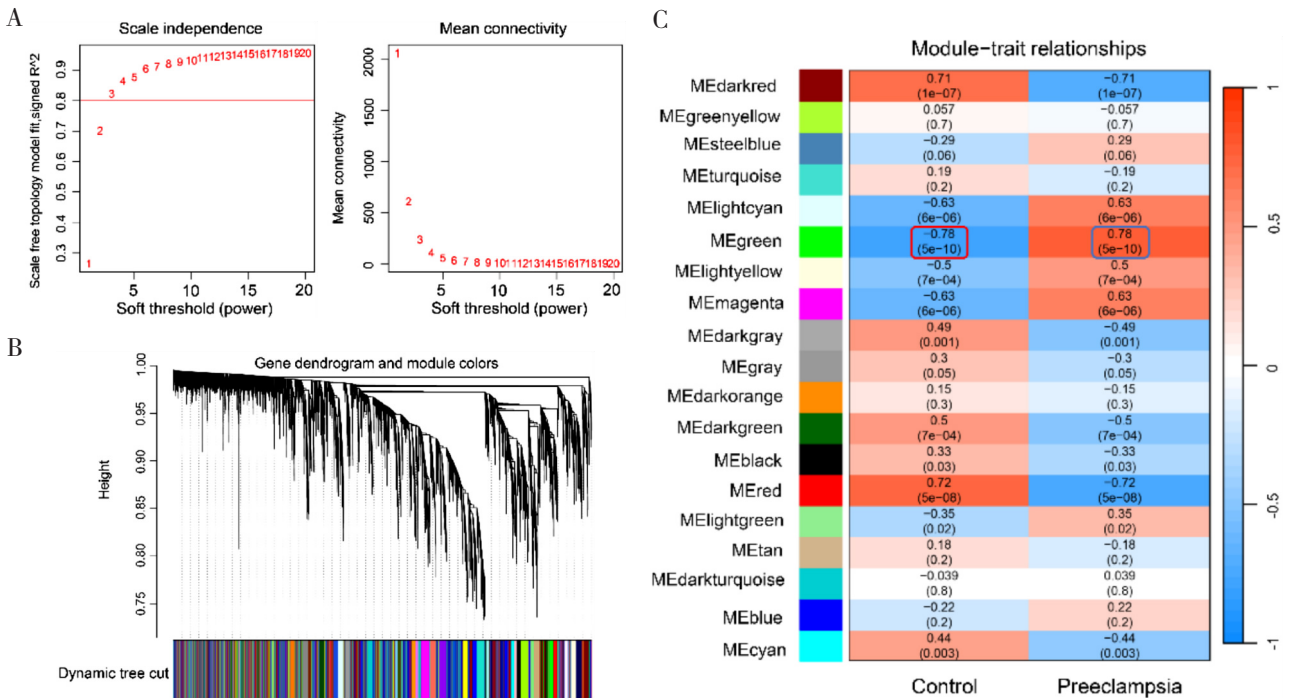


图 1 PE 患者组与对照组差异表达基因图

### 2.2 WGCNA 网络的构建及 PE 相关基因模块识别

为了确定潜在基因模块是否与 PE 相关,对 GSE10588 数据集中的所有 11 338 个基因进行了 WGCNA。选择最佳软阈值为 4,最终确定了 19 个不同的基因模块,经过分析得出了模块中基因与 PE 的相关系数,可以观察到绿色模块的基因与 PE 相关性最大,见图 2。



注:A:最佳软阈值拟合分析;B:所有基因进行的树形聚类产生基因模块图;C:聚类基因模块与 PE 相关性的模块-性状热图

图 2 加权基因共表达网络的构建

### 2.3 重点 DEGs 的鉴定

对 DEGs 集团与 WGCNA 中绿色模块中的基因集团取交集后,总共获得 23 个重点 DEGs,这些基因可能在 PE 疾病的发生发展过程中起重要作用,见图 3。

### 2.4 GO 和 KEGG 通路富集分析

GO 注释分析由分子功能(biological process, BP)、细胞组分(cellular component, CC)和分子功能(molecular function, MF)3 个部分组成。CC 分析结果显示靠前的有:微纤维、细胞桥粒、细胞内不溶性

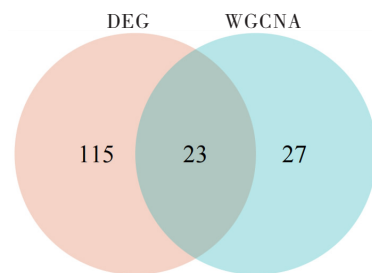


图 3 差异表达基因集团与 WGCNA 中绿色模块基因集团的韦恩图

膜、肌球蛋白复合及黏着斑。BP 中排名靠前的有:调节促性腺激素分泌、白细胞分化、造血功能、内分泌功能及 B 细胞的激活。MF 分析结果显示,激活素的结合、激素活性、跨膜受体蛋白激酶活性、生长因子结合及受体配体的活性等排名靠前。KEGG 富集分析显示,重点基因主要影响细胞因子与细胞因子受体相互作用、溶酶体信号通路、癌症的转录失调和其他聚糖降解通路,见图 4。

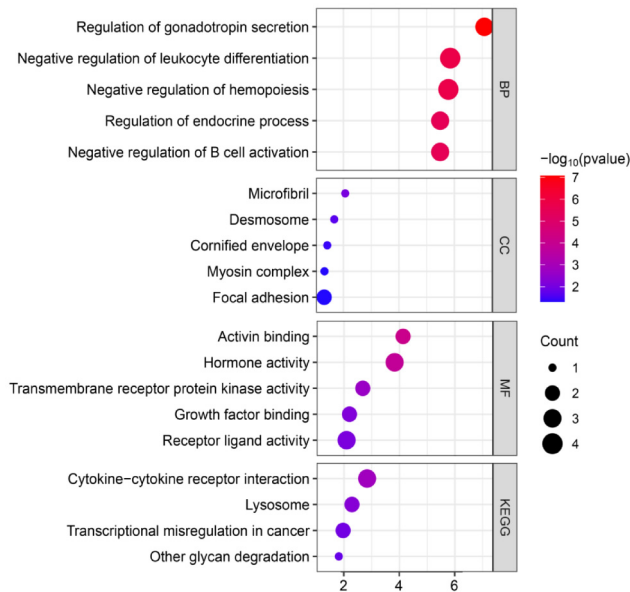
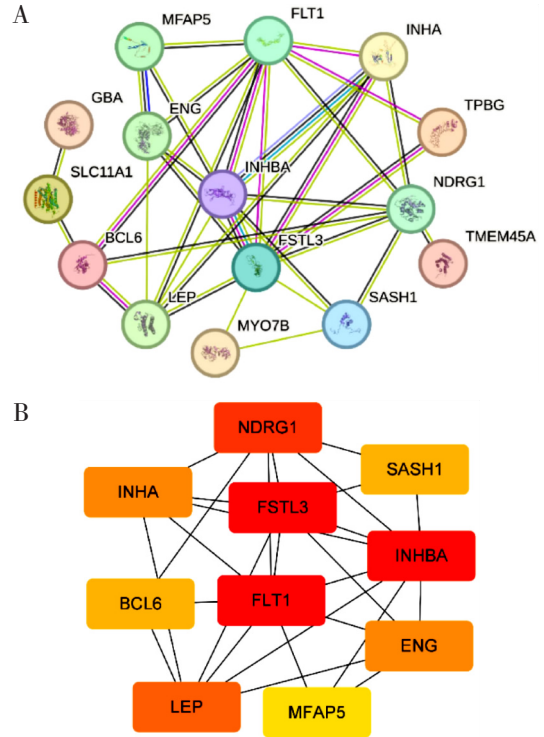


图 4 GO 和 KEGG 通路富集分析

### 2.5 枢纽基因的 PPI 网络分析

通过 STRING 在线工具构建了重点基因的 PPI 网络,见图 5A。随后,使用 Cytoscape 软件对节点最多的前 10 个基因进行可视化,节点最多的基因分别是卵泡抑素相关基因 3 (follistatin-related gene 3, FSTL3)、酪氨酸激酶受体 1 (serum tyrosine kinase receptor 1, FLT1)、抑制素亚单位 A (inhibin subunit beta A, INHBA)、N-myc 下游调节基因 1 (N-myc downstream regulated gene 1, NDRG1)、瘦素 (leptin, LEP)、抑制素  $\alpha$  (inhibin  $\alpha$ , INHA)、内皮糖蛋白

(endothelial glycoprotein, ENG)、SAM 和 SH3 域 1 (SAM-and SH3-domain containing1, SASH1)、B 细胞淋巴瘤 6 (B-cell lymphoma 6, BCL6) 及微原纤维相关蛋白 5 (microfibril-associated protein 5, MFAP5), 见图 5B。

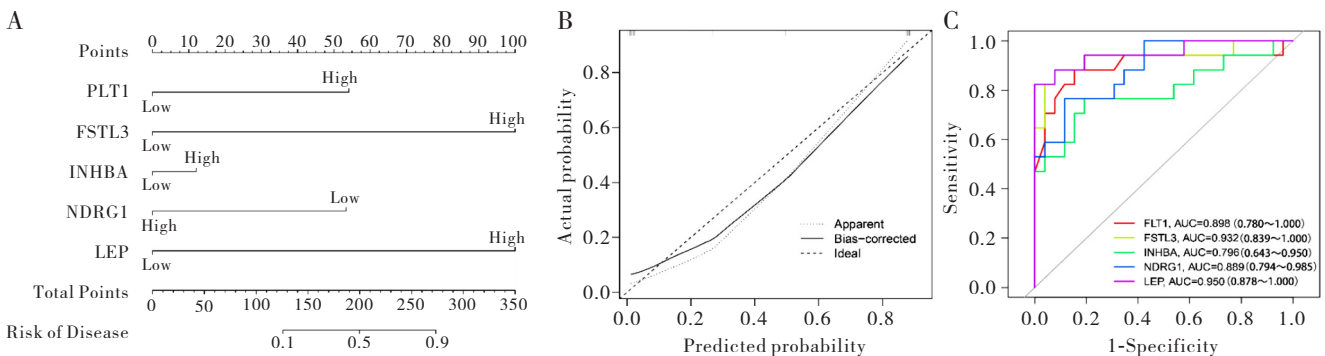


注:A:重点基因的 PPI 网络图; B:节点最多的前 10 个枢纽基因的蛋白-蛋白相互作用网络图

图 5 蛋白-蛋白相互作用网络图

### 2.6 PE 的风险预测及 ROC 曲线

选择前 5 个枢纽基因,构建了风险预测模型。发现 FSTL3 和 LEP 的高表达具有较高的风险得分,见图 6A,校准曲线显示列线图预测和实际结果之间具有较高的一致性,见图 6B。随后,我们计算了这五个枢纽基因的 ROC 曲线,以评估诊断的准确性,见图 6C。结果显示,FSTL3 和 LEP 基因的 AUC 值分别为 0.932 和 0.950,说明具有良好的诊断价值。



注:A:枢纽基因的风险预测列线图; B:预测列线图的校准曲线图; C:ROC 曲线

图 6 PE 的风险预测及诊断评估

### 3 讨论

WGCNA作为一种系统的生物学研究方法,通常用于表型-基因表达之间的遗传关联模式分析,可有效识别候选基因<sup>[13]</sup>。鉴于此,本研究利用WGCNA和PPI网络分析技术,发现了参与PE患者发生发展的枢纽基因和分子途径,旨在探索PE的病理生理机制,为临床治疗提供新的预测、诊断和治疗靶点。

在本研究中,共筛选了138个DEGs,包括100个上调和38个下调的DEGs。WGCNA分析发现,所有DEGs中有23个基因与PE存在密切的关系。GO富集分析显示,这些DEGs主要与激素分泌调节、白细胞分化的负调控、造血的负调控、跨膜受体蛋白激酶活性和生长因子的结合等生物学功能相关。而KEGG通路富集分析显示,这些DEGs主要影响细胞因子与细胞因子受体相互作用、溶酶体以及聚糖降解通路。PPI网络筛选出FSTL3、FLT1、INHBA、NDRG1及LEP五个枢纽基因,最后,经风险预测及ROC曲线分析,确定FSTL3和LEP两个基因在PE中具有较高的风险分数和诊断价值,可作为PE预测及诊断潜在的生物标志物。

FSTL3是一种与转化生长因子 $\beta$ (transforming growth factor  $\beta$ , TGF- $\beta$ )家族结合的蛋白,主要作用是抑制TGF- $\beta$ 家族成员(包括激活素和肌肉生长抑制素)的转录活性<sup>[14]</sup>。近年来研究发现<sup>[15]</sup>,FSTL3在调控多种生物过程中发挥重要作用,包括细胞分化、衰老、肥胖、动脉硬化及肿瘤进展等。FSTL3在人体多种组织中表达,以胎盘中的表达最高<sup>[16]</sup>。激活素A是TGF- $\beta$ 超家族的成员,主要存在于子宫内组织膜中,可促进滋养细胞增殖<sup>[17]</sup>。正常孕妇体内激活素A水平在怀孕后逐渐升高,8至16周下降,在妊娠晚期再次逐渐升高,足月时达到最大值,与妊娠早期相比,足月妊娠母体血清内激活素A增加了7.8倍<sup>[18]</sup>。FSTL3蛋白作为细胞外调节蛋白,可结合激活素A,调控其活性。研究发现<sup>[19]</sup>,FSTL3与激活素A结合后,中和并抑制其功能,当激活素A的表达增加时,FSTL3的表达也会随之增加;注射FSTL3 siRNA混合物的子宫角胚胎着床数量明显少于对照组。这些研究说明FSTL3在正常妊娠的建立和维持中发挥重要作用,激活素A通路的紊乱与PE有关,并且母体血清中的FSTL3水平可以预测PE的发生<sup>[16]</sup>。胎盘缺氧是PE患者的特征之一,研究显示<sup>[20]</sup>,将足月胎盘分离的原代胎盘滋养细胞暴露于缺氧条件4小时后,FSTL3的表达增加了4~6倍。本研究发现FSTL3参与PE发病机制,当FSTL3呈

高表达时,具有较高的风险得分,说明FSTL3极有可能成为预测PE的生物标志物。

LEP是由肥胖基因在7号染色体上编码的蛋白质产物,它由167个氨基酸组成,分子量为16 kDa。LEP在调节脂肪组织能量平衡和分泌方面发挥着重要作用<sup>[21]</sup>。此外,LEP在生殖成熟和生育方面也具有多种调控作用,包括促进滋养细胞增殖、调控卵巢功能和胚胎着床过程。<sup>[22]</sup>。在妊娠期间,母亲体内的LEP处于较高水平,其血清LEP浓度比非孕妇高2~3倍,在妊娠28~32周左右达到峰值,分娩后其水平迅速下降至孕前水平<sup>[23]</sup>。这种显著的LEP增加主要是由脂肪组织和胎盘产生的,与妊娠期间母体体重增加和脂肪沉积有关。目前,胎盘LEP在母体和胎儿中的作用尚未完全阐明,但普遍认为胎盘LEP在调节妊娠相关功能方面起着重要作用<sup>[24]</sup>。研究显示<sup>[25]</sup>,PE孕妇的胎盘中LEP基因附近的CpG位点发生了低甲基化改变,这可能与PE的发生有显著关系。在本研究中,高表达LEP时,具有较高的风险得分,说明LEP也可能是预测PE的生物标志物。

综上所述,本研究结果表明FSTL3和LEP与PE患者的发生发展密切相关,可以作为预测和诊断PE的有效生物标志物。然而,本研究存在一定的不足之处,首先,研究仅在转录组测序数据中进行了分析和预测,没有进行体内、体外或细胞试验验证。其次,样本量较少,这可能影响到结果的可靠性。因此,未来希望有更大样本量的相关研究开展,以进一步提高研究质量获得更可靠的结果。

#### 参考文献:

- [1] MacDonald T M, Walker S P, Hannan N J, et al. Clinical tools and biomarkers to predict preeclampsia[J]. EBioMedicine, 2022, 75: 103780.
- [2] Miller E C, Wilczek A, Bello N A, et al. Pregnancy, preeclampsia and maternal aging: from epidemiology to functional genomics[J]. Ageing Res Rev, 2022, 73: 101535.
- [3] Zeisler H, Llorba E, Chantraine F, et al. Predictive value of the sFlt-1: PIGF ratio in women with suspected preeclampsia[J]. N Engl J Med, 2016, 374(1): 13-22.
- [4] Carbone I F, Conforti A, Picarelli S, et al. Circulating nucleic acids in maternal plasma and serum in pregnancy complications: are they really useful in clinical practice? A systematic review[J]. Mol Diagn Ther, 2020, 24(4): 409-431.
- [5] Xu M M, Zhou H, Hu P, et al. Identification and vali-

- dation of immune and oxidative stress-related diagnostic markers for diabetic nephropathy by WGCNA and machine learning[J]. *Front Immunol*, 2023, 14: 1084531.
- [6] Zeng J F, Lai C, Luo J W, et al. Functional investigation and two-sample Mendelian randomization study of neuropathic pain hub genes obtained by WGCNA analysis[J]. *Front Neurosci*, 2023, 17: 1134330.
- [7] Zheng Y, Gao W Q, Zhang Q, et al. Ferroptosis and autophagy-related genes in the pathogenesis of ischemic cardiomyopathy[J]. *Front Cardiovasc Med*, 2022, 9: 906753.
- [8] Chen S S, Zhang Y C, Ding X Y, et al. Identification of lncRNA/circRNA-miRNA-mRNA ceRNA network as biomarkers for hepatocellular carcinoma[J]. *Front Genet*, 2022, 13: 838869.
- [9] Szklarczyk D, Kirsch R, Koutrouli M, et al. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest [J]. *Nucleic Acids Res*, 2023, 51(D1): D638-D646.
- [10] Zhang D, Hu J T, Liu Z J, et al. Prognostic nomogram in patients with epithelioid sarcoma: a SEER-based study[J]. *Cancer Med*, 2023, 12(3): 3079-3088.
- [11] Li W B, Zhang C, Li W L, et al. Nomogram for predicting fulminant necrotizing enterocolitis[J]. *Pediatr Surg Int*, 2023, 39(1): 154.
- [12] Vélez-Páez J L, Baldeón-Rojas L, Cañadas Herrera C, et al. Receiver operating characteristic (ROC) to determine cut-off points of clinical and biomolecular markers to discriminate mortality in severe COVID-19 living at high altitude[J]. *BMC Pulm Med*, 2023, 23(1): 393.
- [13] Zhao Z Y, He S J, Yu X C, et al. Analysis and experimental validation of rheumatoid arthritis innate immunity gene CYFIP2 and pan-cancer[J]. *Front Immunol*, 2022, 13: 954848.
- [14] Liu Y J, Li J P, Zhang Y, et al. *FSTL3* is a prognostic biomarker in gastric cancer and is correlated with M2 macrophage infiltration[J]. *Onco Targets Ther*, 2021, 14: 4099-4117.
- [15] Duan W F, Shi R X, Yang F, et al. *FSTL3* partially mediates the association of increased nonalcoholic fatty liver disease fibrosis risk with acute myocardial infarction in patients with type 2 diabetes mellitus[J]. *Cardiovasc Diabetol*, 2023, 22(1): 297.
- [16] Gong S, Gaccioli F, Dopierala J, et al. The RNA landscape of the human placenta in health and disease[J]. *Nat Commun*, 2021, 12(1): 2639.
- [17] Zhao H J, Chang H M, Zhu H, et al. Bone morphogenetic protein 2 promotes human trophoblast cell invasion by inducing activin A production [J]. *Endocrinology*, 2018, 159(7): 2815-2825.
- [18] Guo J, Tian T, Lu D H, et al. Alterations of maternal serum and placental follistatin-like 3 and myostatin in pre-eclampsia[J]. *J Obstet Gynaecol Res*, 2012, 38(7): 988-996.
- [19] Xu Y Q, Xie J M, Wan L X, et al. Follistatin-like 3, an activin A binding protein, is involved in early pregnancy loss[J]. *Biomed Pharmacother*, 2020, 121: 109577.
- [20] Aisagbonhi O, Bui T, Nasamran C A, et al. High placental expression of *FLT1*, *LEP*, *PHYHIP* and *IL3RA* in persons of African ancestry with severe preeclampsia [J]. *Placenta*, 2023, 144: 13-22.
- [21] Raskiliene A, Smalinskiene A, Kriaucioniene V, et al. Associations of *MC4R*, *LEP*, and *LEPR* polymorphisms with obesity-related parameters in childhood and adulthood[J]. *Genes*, 2021, 12(6): 949.
- [22] Kapustin R V, Kopteeva E V, Alekseenkova E N, et al. Placental expression of endoglin, placental growth factor, leptin, and hypoxia-inducible factor-1 in diabetic pregnancy and pre-eclampsia[J]. *Gynecol Endocrinol*, 2021, 37(sup1): 35-39.
- [23] Hao S Y, You J, Chen L, et al. Changes in pregnancy-related serum biomarkers early in gestation are associated with later development of preeclampsia [J]. *PLoS One*, 2020, 15(3): e0230000.
- [24] de Kneegt V E, Hedley P L, Kanters J K, et al. The role of leptin in fetal growth during pre-eclampsia[J]. *Int J Mol Sci*, 2021, 22(9): 4569.
- [25] Xiang Y Q, Cheng Y, Li X T, et al. Up-regulated expression and aberrant DNA methylation of *LEP* and *SH3PXD2A* in pre-eclampsia[J]. *PLoS One*, 2013, 8(3): e59753.

[收稿日期 2023-12-13]