Received: 26 September 2024 | Revised: 23 January 2025 | Accepted: 3 February 2025 | Published online: 21 February 2025

DOI: 10.1002/ijgo.70014

CLINICAL ARTICLE

Gynecology

Identification of crucial genes for polycystic ovary syndrome and atherosclerosis through comprehensive bioinformatics analysis and machine learning

Lirong Wang[1] | Yanli Zhang[1] | Fan Ji[1] | Zhenmin Si[1] | Chengdong Liu[2] | Xiaoke Wu[3,4] | Chichiu Wang[5] | Hui Chang[1,3]

[1]The First Clinical Medical College, Heilongjiang University of Chinese Medicine, Harbin, China

[2]Department of Traditional Chinese Medicine, Affiliated Hospital of Jiangsu University, Zhenjiang, China

[3]Department of Obstetrics and Gynecology, First Affiliated Hospital of Heilongjiang University of Chinese Medicine, Harbin, China

[4]Department of Obstetrics and Gynecology, Heilongjiang Provincial Hospital, Harbin, China

[5]Department of Obstetrics and Gynecology, The Chinese University of Hong Kong, Hong Kong, China

Correspondence
Hui Chang, The First Clinical Medical College, Heilongjiang University of Chinese Medicine, No. 24 Heping Road, Wenzheng Street, Harbin, China.
Email: changhui1903@sina.com

Funding information
National Natural Science Foundation of China, Grant/Award Number: 82004404; The Project of Heilongjiang Province "ChunYan" Innovation Team, Grant/Award Number: 2022CYQN0020; Natural Science Foundation of Heilongjiang Province, Grant/Award Number: JJ2024LH0242

Abstract

Objective: To identify potential biomarkers in patients with polycystic ovary syndrome (PCOS) and atherosclerosis, and to explore the common pathologic mechanisms between these two diseases in response to the increased risk of cardiovascular diseases in patients with PCOS.

Methods: PCOS and atherosclerosis data sets were downloaded from the GEO database, and their differentially expressed genes were identified. Weighted gene co-expression network analysis was used to obtain intersection genes, and then protein–protein interaction and functional enrichment analysis were performed. Machine learning algorithms were used to select the key genes, which were then validated through external data sets. We constructed a nomogram to predict the risk of atherosclerosis in women with PCOS. Finally, the CIBERSORT algorithm was used to analyze the infiltration of immune cells in the atherosclerosis group.

Results: We identified six hub genes (CD163, LAPTM5, TNFSF13B, MS4A4A, FGR, and IRF1) that exhibited excellent diagnostic value in validation data sets. Gene ontology terms and KEGG signaling pathway analysis revealed that key genes were associated with immune responses and inflammatory reactions. Abnormal immune cell infiltration was also found in the atherosclerosis group and was correlated with the six hub genes.

Conclusion: Common therapeutic targets of PCOS and atherosclerosis were preliminarily identified through bioinformatics analysis and machine learning techniques. These findings provide new treatment ideas for reducing the risk that PCOS will develop into atherosclerosis.

KEYWORDS
atherosclerosis, biomarkers, machine learning, polycystic ovary syndrome

1 | INTRODUCTION

Among women of reproductive age, polycystic ovary syndrome (PCOS) is the most common endocrine and metabolic disorder. Its principal etiologies are associated with insulin resistance, low-grade chronic inflammation, and elevated androgen levels.[1,2] The diagnosis of PCOS is based on criteria that include hyperandrogenism, ovulatory dysfunction, and polycystic ovary morphology.[3] PCOS not only impacts fertility but is also associated with a range of long-term health problems, such as hypertension, diabetes, and cardiovascular diseases.[4,5] Atherosclerosis is a chronic inflammatory vascular disease, and is the primary pathologic basis for cardiovascular diseases.[6] Studies have shown that women with PCOS are at a significantly greater risk and have a higher incidence of cardiovascular diseases than those without PCOS.[7–9] Moreover, the incidence of acute myocardial infarction among young PCOS patients has been increasing.[10] A large retrospective cohort study investigated adverse cardiovascular events in 174 660 women with PCOS and showed that young women with PCOS had a greater risk of myocardial infarction, angina, and revascularization.[11] Therefore, early identification and intervention are essential for the prevention of cardiovascular diseases in individuals with PCOS.

Chronic inflammation plays a pivotal role in the pathogenesis of both PCOS and atherosclerosis. During the progression of atherosclerosis, the NLRP3 (NOD-,LRR-and pyrin domain-containing 3) inflammasome promotes the release of the proinflammatory cytokine interleukin-1 (IL-1), which enhances inflammatory responses and contributes to thrombosis and vascular occlusion.[12] Studies have shown that elevated levels of inflammatory markers such as C-reactive protein, interleukin-18, tumor necrosis factor-α (TNF-α), interleukin-6, white blood cells, monocyte chemoattractant protein-1, and macrophage inflammatory protein-1α,are closely linked to the pathogenesis of PCOS and may function as early indicators of the risk of atherosclerosis.[13,14] Moreover, hormonal imbalances and chronic low-grade inflammation have been demonstrated to accelerate atherosclerotic processes and the progression of cardiovascular diseases in young PCOS patients.[15] Insulin resistance can impair endothelial function by inhibiting the production of nitric oxide and activating the mitogen-activated protein kinase (MAPK) pathway, facilitating lipid deposition and the infiltration of inflammatory cells.[16] In addition, oxidative stress can also damage vascular endothelial cells and accelerate the progression of atherosclerosis.[17]

PCOS and atherosclerosis are two interrelated chronic conditions that impact patients' cardiovascular health. This study aims to explore the potential molecular mechanisms of these two diseases through bioinformatics analysis and machine learning techniques, providing new perspectives and strategies for the prevention and treatment of PCOS.

2 | MATERIALS AND METHODS

2.1 | Data collection and processing

The PCOS and atherosclerosis data sets were downloaded from the Gene Expression Omnibus (GEO) database.[18] Data sets for PCOS are GSE10946, GSE34526, and GSE137684, and the data sets for atherosclerosis are GSE28829, GSE100927, and GSE43292. Details of six data sets are presented in Table 1. Additionally, the data sets GSE137684 and GSE43292 were used as external validation data sets. The entire research design workflow is depicted in Figure 1. This study was an analysis of publicly available data and did not involve any new human or animal experiments; as a result, it was exempt from ethical approval.

We used the "sva" package to perform batch effect correction. The "limma" package was used to identify differentially expressed genes (DEGs). In the PCOS data set, DEGs with a *P* value less than 0.05 and a |log2FC (fold change)| greater than 0.585 were screened. For atherosclerosis, the criteria were a *P* value adjusted of less than 0.05 and a |log2 FC| greater than 1. Heatmaps and volcano plots provide visual displays for gene expression analysis.

2.2 | Weighted gene co-expression network analysis

Weighted gene co-expression network analysis (WGCNA) is an important bioinformatics method for revealing the complex associations between gene expression and phenotypes.[19] In this study, we used the R package for WGCNA to construct gene co-expression networks of two diseases. At the beginning of the analysis, a scale-free

TABLE 1 Details of polycystic ovary syndrome and atherosclerosis microarray data.

Disease	GSE series	Samples	Platform	Group
PCOS	GSE34526	7 patients and 3 controls	GPL570	Discovery
	GSE10946	13 patients and 11 controls	GPL570	Discovery
	GSE137684	8 patients and 4 controls	GPL17077	Validation
AS	GSE100927	69 patients and 35 controls	GPL17077	Discovery
	GSE28829	16 patients and 13 controls	GPL570	Discovery
	GSE43292	32 patients and 32 controls	GPL6244	Validation

Abbreviations: AS, atherosclerosis; PCOS, polycystic ovary syndrome.

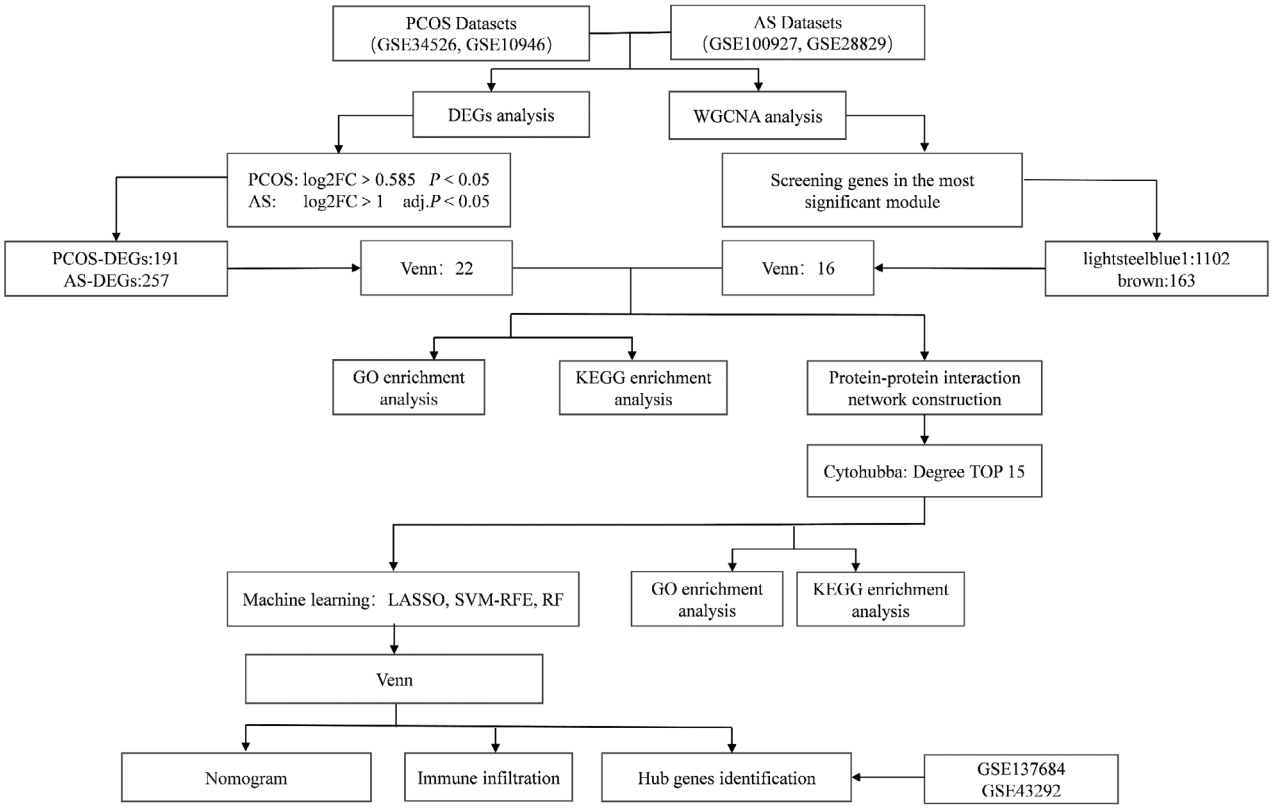

FIGURE 1 Study work flow chart. AS, atherosclerosis; DEGs, differentially expressed genes; GO, gene ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; LASSO, Least absolute shrinkage and selection operator; PCOS, polycystic ovary syndrome; RF, random forest; SVM-RFE, support vector machine-recursive feature elimination; WGCNA, weighted gene co-expression network analysis.

network is constructed and an appropriate β value is selected to form an adjacency matrix, which is transformed into a topologic overlap matrix. Then, the dynamic tree cut algorithm is applied to identify distinct gene modules. Ultimately, we filtered out the gene modules that were most significantly associated with clinical phenotypes.

2.3 | Identification of shared genes and functional enrichment analysis

To explore the biologic functions of these genes and the signaling pathways involved, we used the R package "clusterProfiler" for Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis.[20,21] We selected GO terms and KEGG pathways with *P* values less than 0.05 to analyze relevant biologic processes and molecular functions.

2.4 | Establishment of protein–protein interactions and identification of hub genes

The intersecting genes were input into the STRING database to explore the interactions between protein molecules, specifying *Homo sapiens* as the species and setting confidence score to 0.4.[22] Protein–protein interactions (PPI) network data were visualized via Cytoscape (version V3.9.0), and the top 15 key genes were identified by degree value with the CytoHubba plugin. Additionally, GO and KEGG pathway analyses were conducted on key genes to further explore their biologic functions and the biologic processes.

2.5 | Feature selection via three well-established machine learning algorithms

Three distinct machine learning algorithms (LASSO, SVM, and RF) were used to further identify the key genes.[23] LASSO was used for gene expression data analysis and feature selection. We constructed a regression model using the R package "glmnet" and assessed the model's performance through 10-fold cross-validation. The SVM algorithm was selected for its applicability to small sample data sets. The RF algorithm was applied to classify the significant genes that were filtered out, and the R package "ggplot2" was used for the analysis. The intersection genes acquired from these algorithms were considered potential biomarkers.

2.6 | ROC evaluation and nomogram construction

We used the R package "pROC" to construct receiver operating characteristics (ROC) curves for key genes, and calculated the area

underthecurve(AUC)toevaluatetheirdiagnosticvalueinthevalidation data set. A nomogram was constructed through the R package "rms", and the relative expression levels of each gene were converted into a score. The aggregation of these scores can predict the risk of PCOS progressing to atherosclerosis.

2.7 | Immune cell infiltration

CIBERSORT is an algorithm based on the principle of linear support vector regression, which involves the expression data of 22 types of immune cells and is used to estimate the relative abundance of immune cells.[24] The proportions of 22 types of immune cells in the atherosclerosis group and the control group were analyzed and visualized using the R package "ggplot2", and are presented in bar plots. The differences in the expression of each immune cell type between the two groups were compared using box plots, with *P* values less than 0.05 considered statistically significant. In addition, a heatmap was constructed through the R package "corrplot" to illustrate the correlations between different immune cells during the pathogenesis of atherosclerosis.

3 | RESULTS

3.1 | Identification of DEGs

By analyzing the GSE10946, GSE34526, GSE28829, and GSE100927 data sets, we identified the DEGs for PCOS and atherosclerosis. Preliminary principal component analysis plots revealed significant batch effects in the data from both disease groups (Figure 2a,e). After correction using the sva algorithm, the batch effects between chips were significantly reduced (Figure 2b,f). Volcano plots and heatmaps were subsequently generated to display the DEGs associated with the two diseases. A total of 191 DEGs were identified in the PCOS data sets, including 92 upregulated genes and 99 downregulated genes (Figure 2c). For atherosclerosis, we obtained a total of 257 DEGs, including 206 upregulated genes and 51 downregulated genes (Figure 2g). The DEGs between PCOS and atherosclerosis are illustrated in heatmaps (Figure 2d,h).

3.2 | Screening for key modules via WGCNA

We conducted WGCNA on the GSE10946 and GSE28829 data sets to identify key modules. We initially constructed a gene co-expression network and selected an appropriate soft-thresholding β according to the standard of a scale-free network, setting a correlation coefficient threshold greater than 0.8. In the PCOS group, $\beta=6$ was the optimal "soft" threshold value (Figure 3a). An adjacency matrix was generated using the adjacency function and input into the topologic overlap matrix for hierarchical clustering analysis (Figure 3b). We identified a total of 32 co-expression modules, among which the MElightsteelblue1 module ($R^2=0.47$, $P=0.020$) showed the strongest positive correlation with the PCOS phenotype, containing 164 genes (Figure 3c). For the AS group, $\beta=18$ was the optimal "soft" threshold value (Figure 3d). We identified a total of 12 modules, among which MEcyan ($R^2=-0.59$, $P<0.001$), MEroyalblue ($R^2=0.59$, $P<0.001$), MEsaddlebrown ($R^2=0.67$, $P<0.001$), and MEbrown ($R^2=0.84$, $P<0.001$) demonstrated stronger associations. After screening, we identified MEbrown as a key module, which contains 1102 genes (Figure 3e,f).

3.3 | Analysis of the shared genes and functional enrichment

To thoroughly identify key genes for PCOS and atherosclerosis, a total of 22 genes were obtained by using the online Venn diagram tool (Figure 4a). Similarly, we conducted intersection analysis on the module genes filtered by WGCNA and obtained 16 shared key genes (Figure 4b). Thirty-eight key genes were obtained by combining two methods, and further functional annotation and enrichment analyses were performed to explore their potential functions and roles in both diseases (Figure 4c,d). GO terms showed that 38 genes were significantly enriched in multiple biologic processes, including positive regulation of cytokine production, phagocytosis, the cellular defense response, activation of the innate immune response, and mononuclear cell differentiation. Notably, these biologic processes play crucial roles in immune responses and inflammatory reactions. KEGG signaling pathway analysis revealed seven signaling pathways related to immune responses, inflammatory processes, and lipid metabolism, including osteoclast differentiation, lipid and atherosclerosis, hematopoietic cell lineage, and cytokine-cytokine receptor interaction.

3.4 | PPI network construction and hub gene selection

The analysis showed that the PPI network comprising these genes contained 38 nodes and 83 edges (Figure 5a). We subsequently imported these genes into Cytoscape software and used the Cytohubba plugin to further select the top 15 key genes on the basis of degree ranking, including ITGAM, CD163, HLA-DRA, VSIG4, SIGLEC1, TREM2, HLA-DPA1, LAPTM5, TNFSF13B, CD52, MS4A4A, FGR, STAB1, IRF1, and HLA-DMA (Figure 5b). Additionally, we performed functional annotation and enrichment analysis on these key genes and found that the majority of the biologic processes and enriched pathways were intimately related to immune system functions (Figure 5c,d).

3.5 | Candidate gene identification using machine learning algorithms

On the basis of the above 15 shared genes, we applied three different machine learning algorithms (including LASSO, SVM-RFE

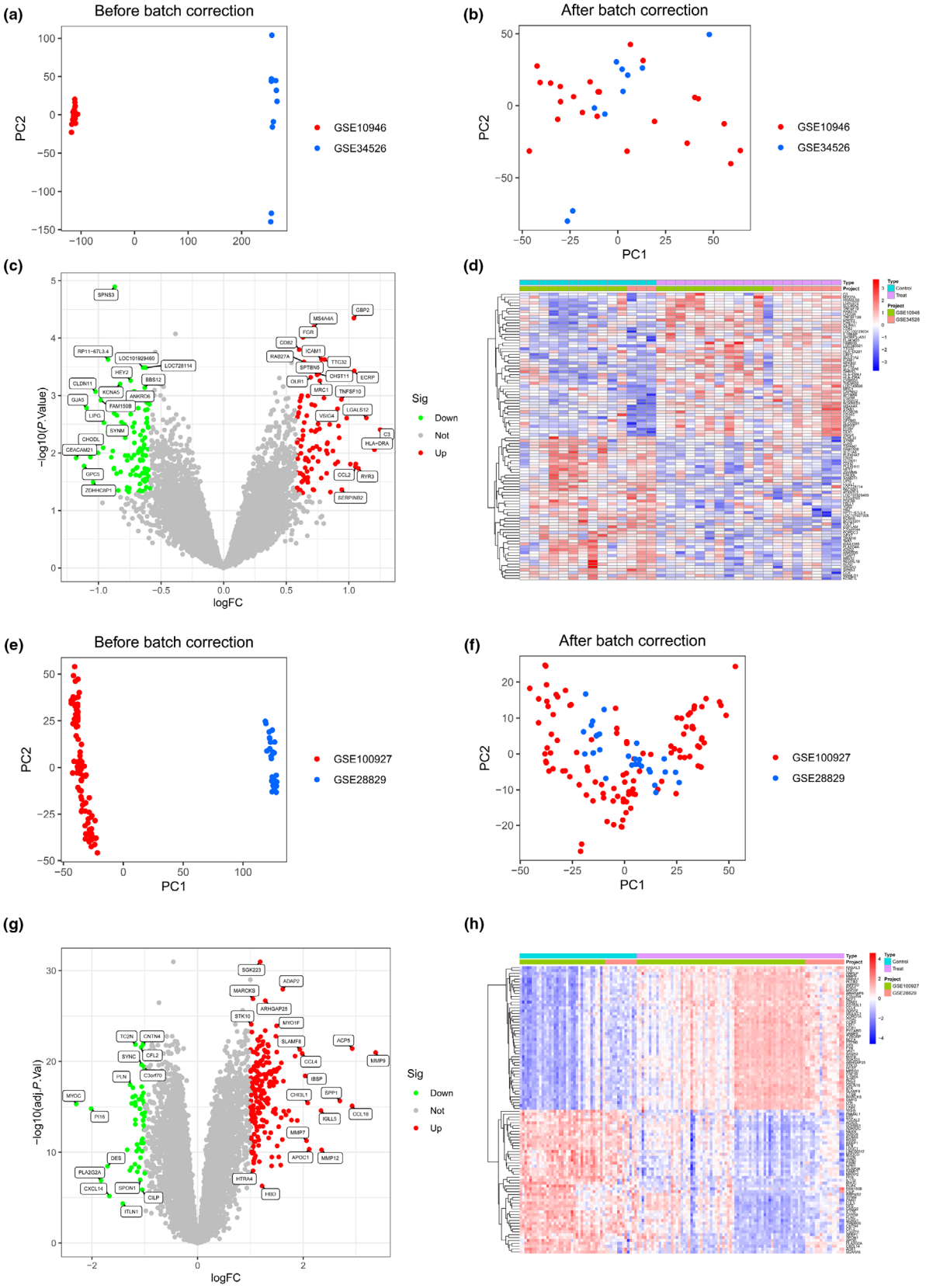

FIGURE 2 Removal of batch effects and identification of differentially expressed genes (DEGs) in polycystic ovary syndrome (PCOS) and atherosclerosis. (a, b) Principle component analysis (PCA) plots showed the expression pattern in two data sets of PCOS before and after eliminating the batch effects. (c, d) DEG volcano and heatmap plot in PCOS group. (e, f) PCA plots showed the expression pattern in two data sets of atherosclerosis before and after removing the batch effects. (g, h) Volcano and the heatmap plot of DEGs in atherosclerosis group.

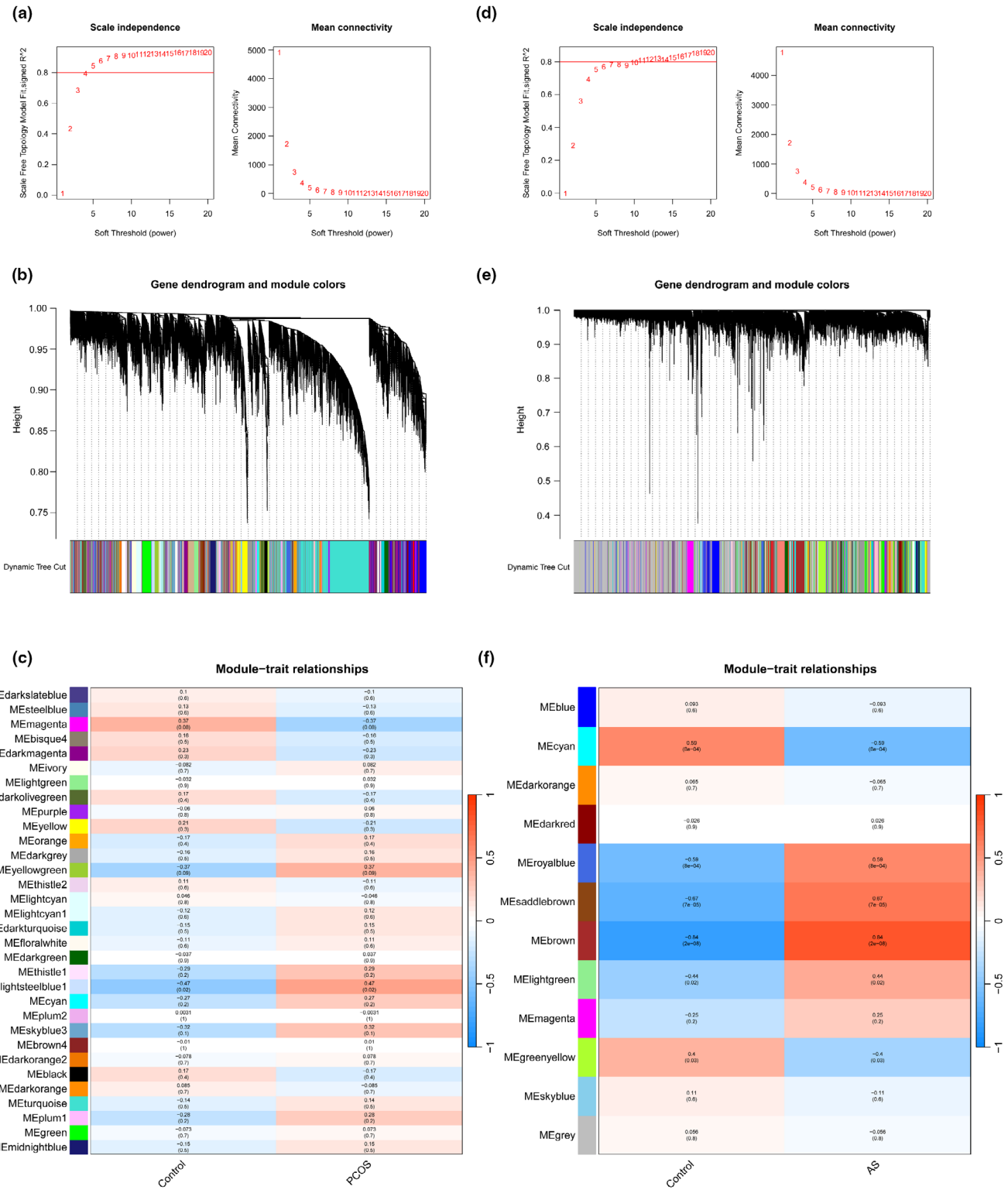

FIGURE 3 Weighted gene co-expression network analysis of polycystic ovary syndrome (PCOS) and atherosclerosis. (a) Determination of soft-threshold power for PCOS. (b) Cluster dendrogram of PCOS highly connected genes in key modules. (c) Relationships between modules and traits in PCOS. Correlations and *p* values are included in each cell. (d) Calculation of soft-threshold power for atherosclerosis. (e) Cluster dendrogram of atherosclerosis modules with highly connected genes. (f) Module-trait relationships in atherosclerosis. A correlation and *p* value are included in each cell.

and RF) to screen for key genes associated with both diseases. The LASSO regression analysis identified 14 DEGs with the lowest binomial deviation (Figure 6a). The SVM-RFE analysis revealed that 13 biomarkers had the highest accuracy and the smallest error in screening PCOS candidate biomarkers (Figure 6b). The RF algorithm was used to calculate the gene importance scores for DEGs, and

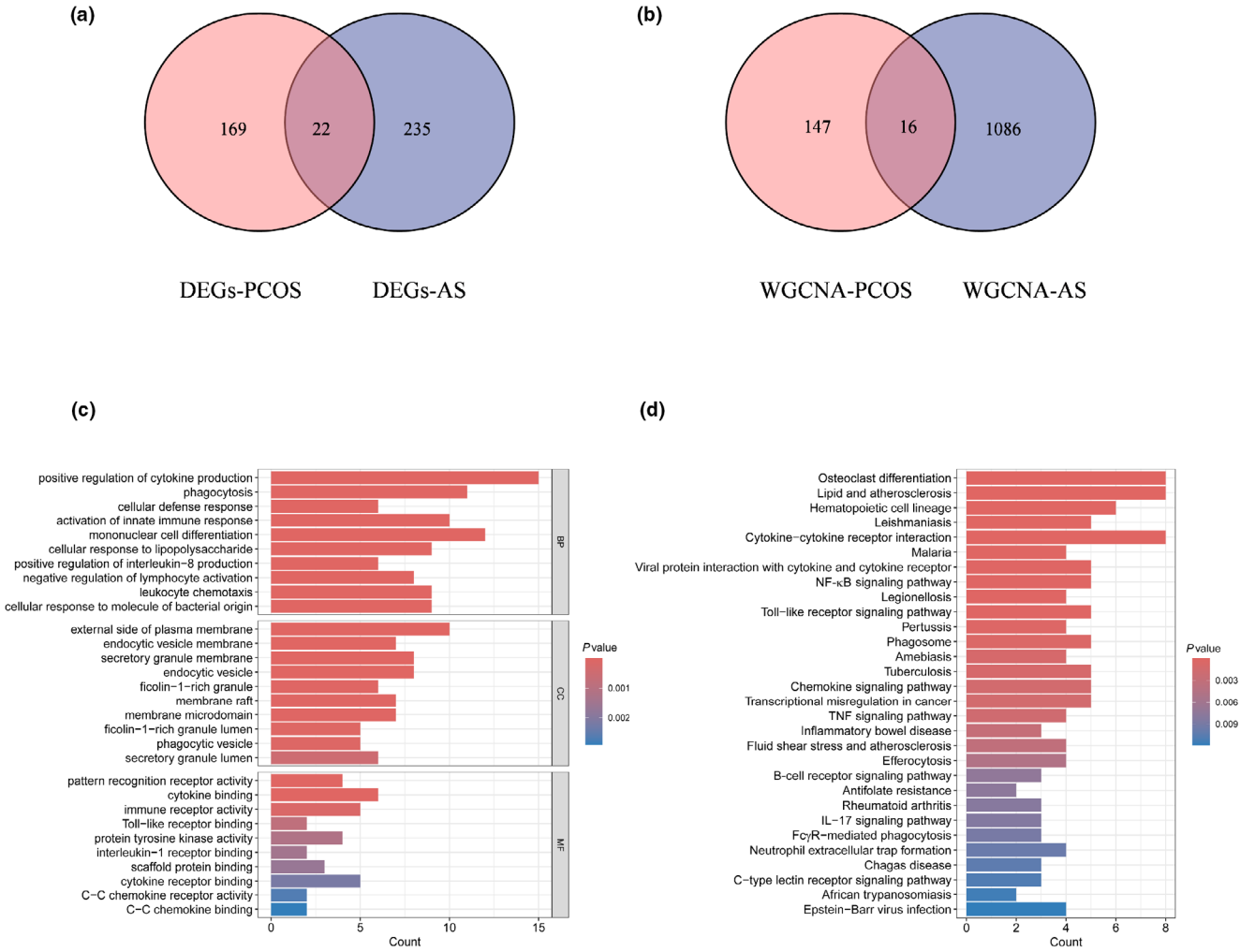

FIGURE 4 Shared gene signatures and functional enrichment between polycystic ovary syndrome (PCOS) and atherosclerosis (AS). (a) Shared differentially expressed genes (DEGs) between PCOS and AS by overlapping their DEGs. (b) Shared genes between the weighted gene co-expression network analysis (WGCNA) modules of PCOS and AS by overlapping them. (c, d) Shared genes were represented by bar plots displaying GO and KEGG enrichment. GO, gene ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes.

genes with scores greater than 1 were selected. Seven genes that fulfilled the screening criteria were identified (Figure 6c). Finally, we analyzed the intersection of the Venn diagrams of the key genes obtained by these three machine learning methods and found that six DEGs, namely, CD163, LAPTM5, TNFSF13B, MS4A4A, FGR, and IRF1, intersected across all methods (Figure 6d). These results suggest that these genes may play a central role in the pathogenesis of both diseases.

3.6 | Evaluation of the predictive value and nomogram construction

To evaluate the specificity and sensitivity of the six key genes for the diagnosis of these two diseases, we performed ROC analysis. In the PCOS group, the following genes had considerable diagnostic value: CD163 (AUC = 0.719, 95% confidence interval [CI] 0.34–1.00), LAPTM5 (AUC = 0.781, 95% CI 0.47–1.00), TNFSF13B (AUC = 0.781, 95% CI 0.47–1.00), MS4A4A (AUC = 0.750, 95% CI 0.31–1.00), FGR (AUC = 0.781, 95% CI 0.50–1.00), and IRF1 (AUC = 0.844, 95% CI 0.56–1.00) (Figure 7a). In the atherosclerosis group, the same ROC analysis was conducted, and each biomarker showed significant predictive performance: CD163 (AUC = 0.850, 95% CI 0.75–0.93), LAPTM5 (AUC = 0.828, 95% CI 0.72–0.92), TNFSF13B (AUC = 0.777, 95% CI 0.66–0.89), MS4A4A (AUC = 0.767, 95% CI 0.64–0.88), FGR (AUC = 0.797, 95% CI 0.69–0.89), and IRF1 (AUC = 0.804, 95% CI 0.69–0.90) (Figure 7b). We observed that the performance of these markers in the atherosclerosis data set was superior to that in the PCOS data set, suggesting that these genes may play a more critical role in the pathophysiologic processes of atherosclerosis. Additionally, a nomogram was constructed to predict the associations between six key genes and the risk of atherosclerosis. In this model, the relative expression level of each gene is assigned a specific score, and the sum of these scores yields a total score that is directly correlated with the risk of atherosclerosis development in individuals with PCOS (Figure 7c).

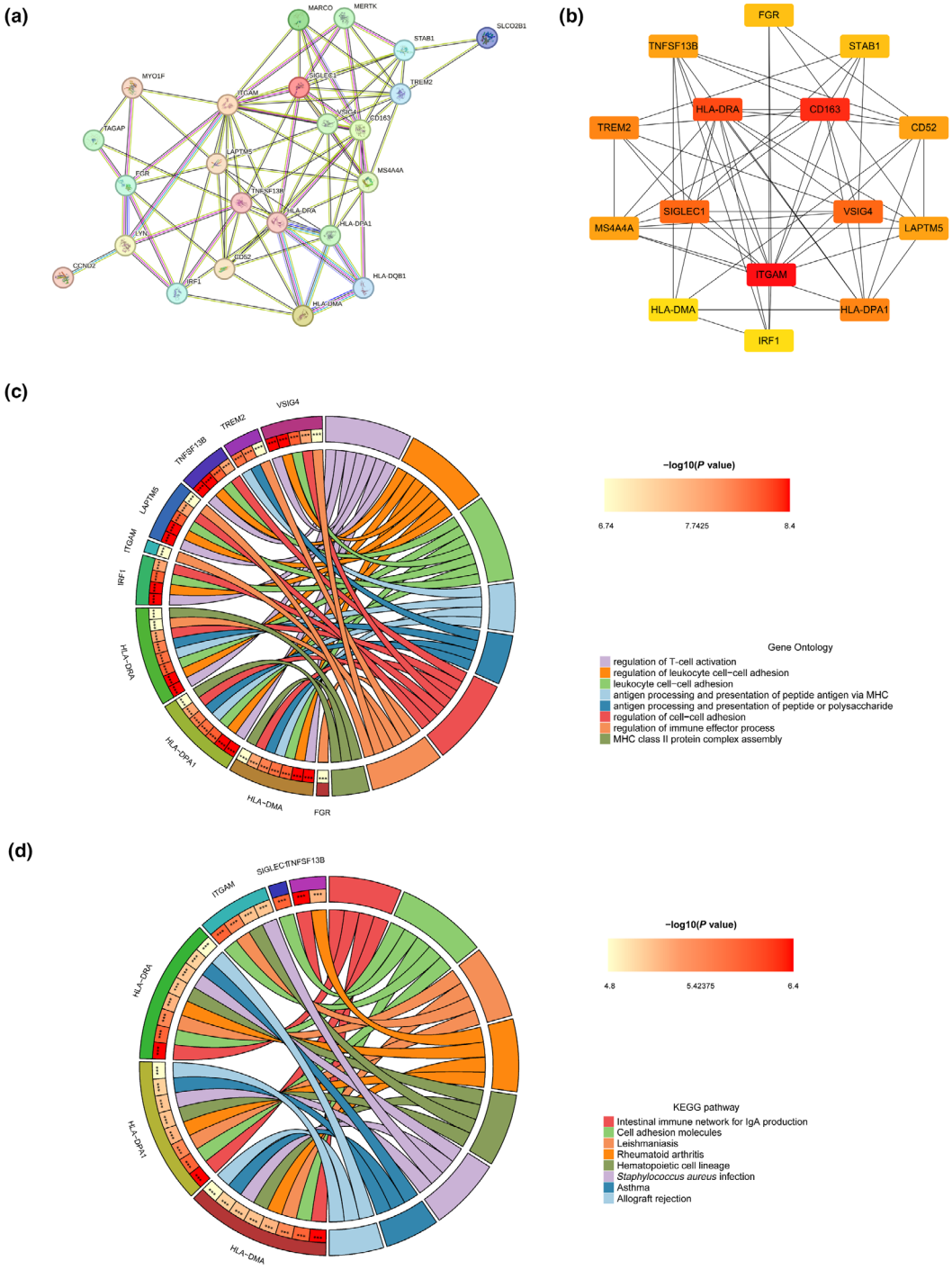

FIGURE 5 Shared gene protein–protein interactions network and functional enrichment of polycystic ovary syndrome and atherosclerosis. (a) Interaction network of proteins using the STRING database. The remaining 15 differentially expressed genes were eliminated due to lack of interaction. (b) Top 15 hub genes of degree by Cytohubba. (c) Gene ontology (GO) analysis (biologic process, cellular component, and molecular function) of 15 common genes. (d) Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis of 15 common genes.

3.7 | Immune infiltration analysis

The above analysis results suggest that atherosclerosis is closely related to the immune response. We used CIBERSORT to analyze the abundance of immune cells in the atherosclerosis data set and visually presented the proportion of immune cells in each sample using a bar chart (Figure 8a). Compared with those in the control group, the abundances of memory B cells, γδ T cells, M0 macrophages, and activated mast cells were greater, whereas the abundances of naive B cells, plasma cells, $CD8^+$ T cells, resting memory $CD4^+$ T cells, monocytes, M1 macrophages, and resting mast cells were lower (Figure 8b). Correlation analysis revealed that activated mast

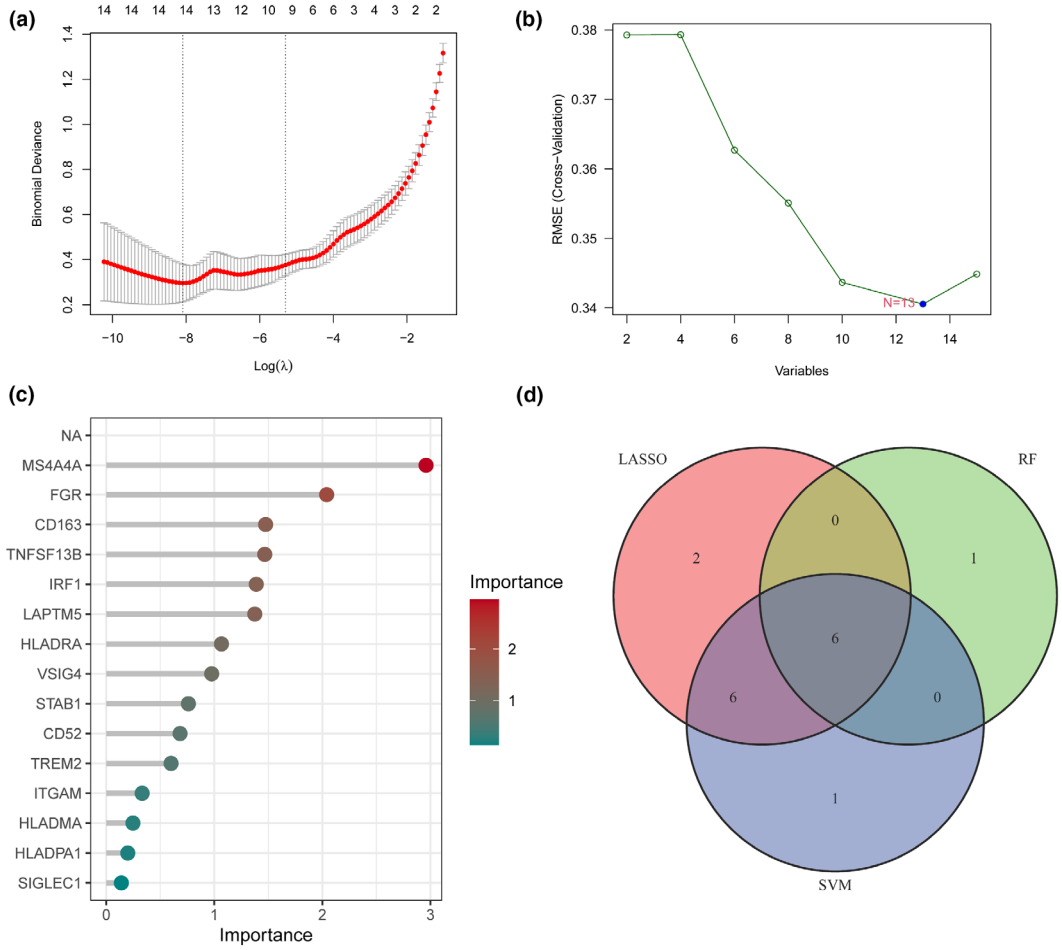

FIGURE 6 Candidate biomarker identification via machine learning algorithms. (a) Based on the Lasso regression algorithm, 14 genes were identified as the biomarkers with the lowest binominal deviation. (b) Thirteen genes were screened by SVM-RFE. (c) Screening genes by importance score based on random forest algorithm. (d) Six hub genes (CD163, LAPTM5, TNFSF13B, MS4A4A, FGR, IRF1) visualize the intersection of three machine learning algorithms by Venn diagram.

cells and resting mast cells had the greatest negative correlation ($r = -0.73$), and the greatest positive correlation was observed between resting memory CD4$^+$ T cells and monocytes ($r = 0.41$) (Figure 8c). Furthermore, relationships were found between biomarkers and the content of immune cells—as resting memory CD4$^+$ T cells and activated natural killer cells were positively correlated with CD163, LAPTM5, TNFSF13B, MS4A4A, and FGR, whereas follicular helper T cells and M0 macrophages were negatively correlated with these biomarkers (Figure 8c).

4 | DISCUSSION

In recent years, the relationship between PCOS and cardiovascular disease has garnered widespread attention. A meta-analysis revealed that PCOS is associated with significant changes in functional and structural markers of subclinical atherosclerosis, suggesting that PCOS can promote the occurrence of subclinical atherosclerosis.[25] Other studies have shown that the hearts of women with PCOS are characterized by increased macrophage accumulation, which significantly exacerbates the development of atherosclerotic plaques.[26] Therefore, finding biomarkers for these two diseases is key to improving the prognosis of PCOS patients. In this study, we identified six key genes from an array of gene data: CD163, LAPTM5, TNFSF13B, MS4A4A, FGR, and IRF1. These genes have good predictive value for promoting the development of atherosclerosis in PCOS and provide new clues for understanding the potential mechanisms of the interaction between PCOS and cardiovascular diseases.

The scavenger receptor cysteine-rich type 1 protein M130 (CD163) is a receptor that is expressed primarily on the surface of tissue macrophages and may play an important role in PCOS and atherosclerosis. One study observed that the percentage of CD163$^+$ M2 macrophages in the endometria of patients with PCOS was abnormally increased, suggesting that CD163$^+$ macrophages may be involved in the immune response and inflammatory process in PCOS.[27] Another study found that the level of soluble CD163 in PCOS patients was significantly increased and positively correlated with total testosterone, total cholesterol, and luteinizing hormone levels.[28] Hence, we speculate that CD163 may be related to the inflammatory response in PCOS. In atherosclerosis, previous research

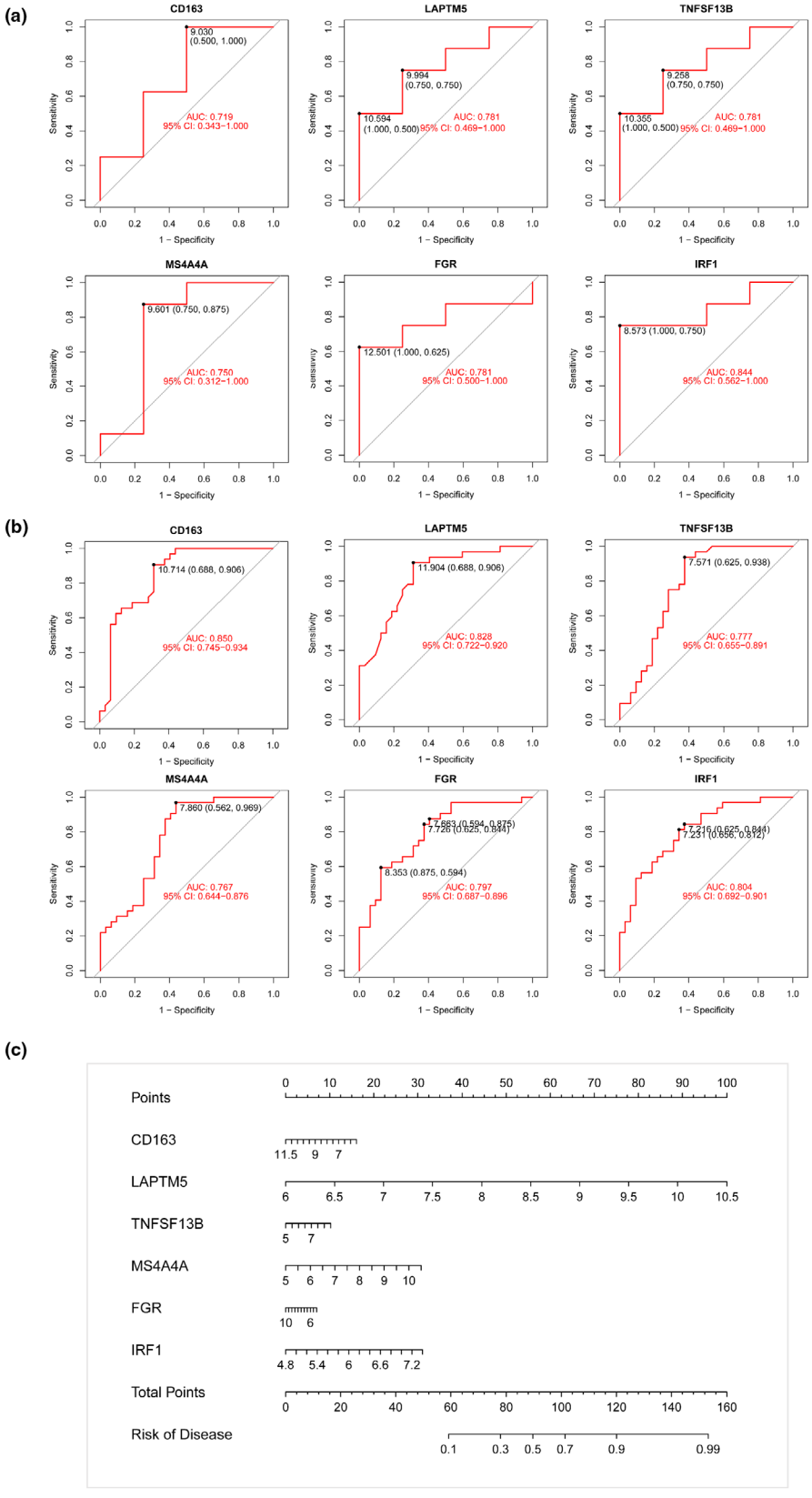

FIGURE 7 Construction of the nomogram and receiver operating characteristics (ROC) curves of the six common core genes in polycystic ovary syndrome (PCOS) and atherosclerosis. (a) The ROC curve of each candidate gene (CD163, LAPTM5, TNFSF13B, MS4A4A, FGR, IRF1) was verified in the PCOS data set GSE137684. (b) The ROC curve of each candidate gene (CD163, LAPTM5, TNFSF13B, MS4A4A, FGR, IRF1) was verified in the atherosclerosis data set GSE43292. (c) Nomogram for predicting the risk of atherosclerosis in PCOS. AUC, area under the curve; CI, confidence interval.

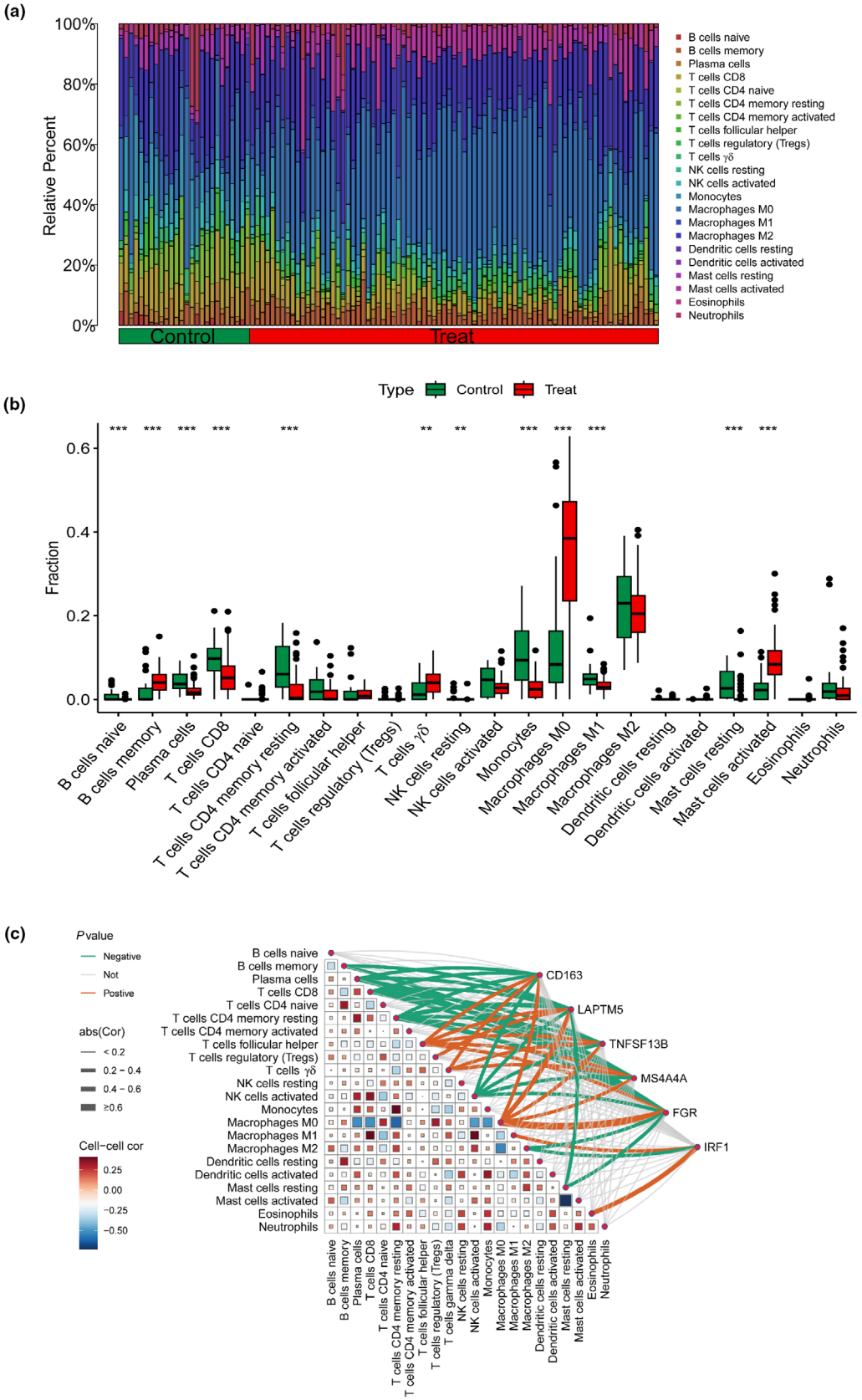

FIGURE 8 Immune infiltration analysis between atherosclerosis and control and correlations between the hub differentially expressed genes (DEGs) and immunologic features in atherosclerosis. (a) Barplot shows the proportion of immune cells in different samples. (b) Boxplot compares the expression of immune cells between atherosclerosis and controls. $^{*}P<0.05$; $^{**}P<0.01$; $^{***}P<0.001$. (c) Correlation analysis of immune cell infiltrations with six hub DEGs.

has shown that CD163 knockout increases atherosclerotic lesion size, and it is regarded as an atheroprotective.[29] However, an increasing number of studies have shown that CD163 has a proatherosclerotic effect. The pathogenic role of $CD163^+$ macrophages is closely related to their role in promoting angiogenesis, vascular permeability, and plaque progression. Specifically, macrophages activate HIF1α by inhibiting prolyl hydroxylase, thereby promoting VEGF-mediated angiogenesis, vascular permeability, and the recruitment of inflammatory cells within the plaque.[30] A study demonstrated that the expression levels of CD163 mRNA and protein are increased in the plaques of patients with cerebrovascular symptoms.[31] Another study further revealed that $CD163^+$ macrophages inhibit vascular calcification through nuclear factor κB (NF-κB)-induced enhancement of hyaluronan synthase, thus promoting the development of high-risk plaques.[32]

Lysosomal-associated transmembrane protein 5 (LAPTM5) has been demonstrated to be involved in the regulation of immune responses, inflammation, apoptosis,[33] and autophagy.[34–36] As a positive regulatory factor in proinflammatory pathways in macrophages, LAPTM5 plays a necessary role in the activation of the NF-κB and MAPK signaling pathways mediated by TNF receptors.[37] The mRNA expression levels of LAPTM5 are significantly increased in mice with myocardial infarction and diabetic nephropathy, and when LAPTM5 is downregulated, the inflammatory response can be significantly improved.[38,39] The mechanism by which it participates in the inflammatory response may be related to activation of the RIP1/NF-κB pathway.[40] However, in research on cerebral ischemia–reperfusion injury and nonalcoholic steatohepatitis, LAPTM5 gene knockout intensified inflammatory responses. These findings indicate that LAPTM5 may help to inhibit inflammation under certain conditions.[41,42] Therefore, LAPTM5 is involved in the processes of inflammation and immune regulation, indirectly influencing the pathogenesis of PCOS and atherosclerosis.

Tumor necrosis factor ligand superfamily member 13B (TNFSF13B) is also called B-cell-activating factor (BAFF). TNFSF13B plays a biphasic role in the inflammatory response. In the early stages of inflammation, BAFF binding to its receptors BAFFR, TACI, and BCMA can promote B-cell survival and proliferation, participating in the positive regulation of immune responses.[43] An increase in TNFSF13B levels is associated with autoimmune diseases such as systemic lupus erythematosus and rheumatoid arthritis,[44,45] which suggests that its overexpression promotes the progression of inflammatory responses. Chen et al.[46] demonstrated that TNFSF13 participates in immunosuppression through various immunoregulatory pathways and regarded it as a new diagnostic biomarker and a potential therapeutic target. Furthermore, higher levels of BAFF in patients with myocardial infarction are associated with an increased risk of major adverse cardiovascular events.[47] Considering the function of TNFSF13B in regulating B-cell-mediated immune responses, we speculate that it may be indirectly involved in the inflammatory responses and immune dysregulation of PCOS.

Membrane-spanning 4-domain subfamily A member 4A (MS4A4A) is a member of the transmembrane four-domain family A (MS4A) subfamily. MS4A proteins control cell activation by acting as ion channels or modulating the signal transduction of other immune receptors, including B-cell receptors, other immunoglobulin receptors, pattern recognition receptors, or trigger receptors.[48] MS4A4A is co-expressed on the cell membrane of monocytes as well as in the inflammatory synovium and tumor-associated macrophages.[49,50] MS4A4A promotes M2 polarization of macrophages by activating the PI3K/AKT and JAK/STAT6 signaling pathways, which suggests that MS4A4A is involved in inflammatory and immune responses.[51] In allergic asthma the expression level of MS4A4A is related to the increased expression of cytokines associated with type 2 immune responses, which further demonstrates the role of MS4A4A in regulating allergic immune responses.[52]

Tyrosine-protein kinase Fgr (FGR) is a nonreceptor protein tyrosine kinase expressed in many cell types. FGR positively regulates the activity of mast cells by activating spleen tyrosine kinase, which is involved in the pathogenesis of various allergic diseases.[53] Studies have shown that knockout of the Fgr gene impairs the polarization of proinflammatory macrophages, and that the activation of Fgr kinase by reactive oxygen species can lead to the activation of proinflammatory adipose tissue macrophages.[54] Furthermore, in sepsis-associated encephalopathy, FGR kinase inhibitors have been found to improve mitochondrial dysfunction, oxidative stress, and neuroinflammation.[55] In summary, as an oncogene, FGR can regulate the function of immune cells and the production of inflammatory mediators, directly or indirectly participating in the pathologic processes of PCOS and atherosclerosis.

Interferon regulatory factor 1 (IRF1) is a pivotal transcriptional regulator of immune responses. It plays a significant role in the development, differentiation, and function of immune cells by modulating the expression of various cytokines, interferons, and immune-related genes. Research indicates that IRF1 is a key molecule in the progression of sepsis-induced acute respiratory distress syndrome, promoting the polarization of macrophages to the M1 phenotype, and that the deletion of IRF1 significantly alleviates lipopolysaccharide-induced lung injury and M1 polarization infiltration.[56] In the study of ovarian cancer, IRF1 has been identified as a downstream target of the IL-6 signaling pathway and serves as a regulator of androgen receptor expression, making it a potential therapeutic target for ovarian cancer.[57] Under the activation of the noncanonical NF-κB signaling pathway, the expression of IRF1 is upregulated, which promotes pyroptosis and the development of atherosclerosis.[58] The latest research suggests that chondrocytes lacking IRF1 accumulate irreversible DNA damage under oxidative stress, accelerating cellular senescence and leading to cartilage destruction and osteoarthritis.[59]

In conclusion, the present study employed a variety of bioinformatics analyses and machine learning algorithms to rigorously identify six key genes (CD163, LAPTM5, TNFSF13B, MS4A4A, FGR, and IRF1) and established a nomogram to predict the risk of atherosclerosis in patients with PCOS. These findings provide new

insights into the potential pathogenic mechanisms underlying the relationship between PCOS and atherosclerosis. However, there are several limitations in this study. First, when relevant transcript data were selected from online databases, the chosen case samples lacked more extensive related clinical information. Second, the relatively small sample size of this study may affect the accuracy of the results. Future research is needed to explore the molecular connections between PCOS and atherosclerosis, as well as to investigate how lifestyle changes, pharmacologic treatments, or other interventions can reduce the risk of cardiovascular diseases in patients with PCOS.

AUTHOR CONTRIBUTIONS

LW contributed to the methodology, software, writing the original draft, and editing. YZ contributed to software, data processing, and editing. FJ contributed to data collection and editing. ZS contributed to data processing and editing. CL contributed to guiding the writing and supervision. XW and CW contributed to supervision and funding acquisition. HC contributed to planning the project and controlling its progress, and guiding the writing. All authors have read and approved the final article.

FUNDING INFORMATION

Funding was received from the National Natural Science Foundation of China (Youth Project) (Grant number 82004404); The Project of Heilongjiang Province "ChunYan" Innovation Team (Grant number 2022CYQN0020); and Heilongjiang Provincial Natural Science Foundation (JJ2024LH0242).

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

REFERENCES

1. Xu Y, Qiao J. Association of insulin resistance and elevated androgen levels with polycystic ovarian syndrome (PCOS): a review of literature. *J Healthc Eng*. 2022;2022:9240569.
2. Deng H, Chen Y, Xing J, Zhang N, Xu L. Systematic low-grade chronic inflammation and intrinsic mechanisms in polycystic ovary syndrome. *Front Immunol*. 2024;15:1470283.
3. Revised 2003 Consensus on Diagnostic Criteria and Long-Term Health Risks Related to Polycystic Ovary Syndrome (PCOS). *Hum Reprod (Oxford, England)*. 2004;19(1):41-47.
4. Grigoryan OR, Zhemaite NS, Volevodz NN, Andreeva EN, Melnichenko GA. Dedov II. Long-term consequences of polycystic ovary syndrome. *Ter Arkh*. 2017;89(10):75-79.
5. Anagnostis P, Tarlatzis BC, Kauffman RP. Polycystic ovarian syndrome (PCOS): long-term metabolic consequences. *Metabolism*. 2018;86:33-43.
6. Zhu Y, Xian X, Wang Z, et al. Research progress on the relationship between atherosclerosis and inflammation. *Biomolecules*. 2018;8(3):80.
7. Wekker V, van Dammen L, Koning A, et al. Long-term cardiometabolic disease risk in women with PCOS: a systematic review and meta-analysis. *Hum Reprod Update*. 2020;26(6):942-960.
8. Ollila MM, Arffman RK, Korhonen E, et al. Women with PCOS have an increased risk for cardiovascular disease regardless of diagnostic criteria-a prospective population-based cohort study. *Eur J Endocrinol*. 2023;189(1):96-105.
9. Benham JL, Goldberg A, Teede H, Tay CT. Polycystic ovary syndrome: associations with cardiovascular disease. *Climacteric*. 2024;27(1):47-52.
10. Arora S, Stouffer GA, Kucharska-Newton AM, et al. Twenty year trends and sex differences in young adults hospitalized with acute myocardial infarction. *Circulation*. 2019;139(8):1047-1056.
11. Berni TR, Morgan CL, Rees DA. Women with polycystic ovary syndrome have an increased risk of major cardiovascular events: a population study. *J Clin Endocrinol Metab*. 2021;106(9):e3369-e3380.
12. Grebe A, Hoss F, Latz E. NLRP3 inflammasome and the IL-1 pathway in atherosclerosis. *Circ Res*. 2018;122(12):1722-1740.
13. Rudnicka E, Suchta K, Grymowicz M, et al. Chronic low grade inflammation in pathogenesis of PCOS. *Int J Mol Sci*. 2021;22(7):3789.
14. Marciniak A, Nawrocka Rutkowska J, Brodowska A, Wiśniewska B, Starczewski A. Cardiovascular system diseases in patients with polycystic ovary syndrome—the role of inflammation process in this pathology and possibility of early diagnosis and prevention. *Ann Agric Environ Med*. 2016;23(4):537-541.
15. Alvarez YR, Pico M, Ashokprabhu N, et al. Polycystic ovarian syndrome: a risk factor for cardiovascular disease. *Curr Atheroscler Rep*. 2023;25(12):1003-1011.
16. Di Pino A, DeFronzo RA. Insulin resistance and atherosclerosis: implications for insulin-sensitizing agents. *Endocr Rev*. 2019;40(6):1447-1467.
17. Förstermann U, Xia N, Li H. Roles of vascular oxidative stress and nitric oxide in the pathogenesis of atherosclerosis. *Circ Res*. 2017;120(4):713-735.
18. Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res*. 2013;41(Database issue):D991-D995.
19. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform*. 2008;9:559.
20. The gene ontology resource: enriching a gold mine. *Nucleic Acids Res*. 2021;49(D1):D325-D334.
21. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28(1):27-30.
22. 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. *Nucleic Acids Res*. 2023;51(D1):D638-D646.
23. Uddin S, Khan A, Hossain ME, Moni MA. Comparing different supervised machine learning algorithms for disease prediction. *BMC Med Inform Decis Mak*. 2019;19(1):281.
24. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015;12(5):453-457.
25. Sun D, Wu Y, Ding M, Zhu F. Comprehensive meta-analysis of functional and structural markers of subclinical atherosclerosis in women with polycystic ovary syndrome. *Angiology*. 2022;73(7):622-634.
26. Gao L, Zhao Y, Wu H, et al. Polycystic ovary syndrome fuels cardiovascular inflammation and aggravates ischemic cardiac injury. *Circulation*. 2023;148(24):1958-1973.
27. Liu S, Hong L, Mo M, et al. Evaluation of endometrial immune status of polycystic ovary syndrome. *J Reprod Immunol*. 2021;144:103282.
28. Oruç AS, Yilmaz N, İnal HA, et al. A study of serum soluble CD 163 levels in women with polycystic ovary syndrome. *Horm Metab Res*. 2016;48(6):399-403.

29. Gutiérrez-Muñoz C, Méndez-Barbero N, Svendsen P, et al. CD163 deficiency increases foam cell formation and plaque progression in atherosclerotic mice. *FASEB J*. 2020;34(11):14960-14976.
30. Guo L, Akahori H, Harari E, et al. CD163+ macrophages promote angiogenesis and vascular permeability accompanied by inflammation in atherosclerosis. *J Clin Invest*. 2018;128(3):1106-1124.
31. Bengtsson E, Hultman K, Edsfeldt A, et al. CD163+ macrophages are associated with a vulnerable plaque phenotype in human carotid plaques. *Sci Rep*. 2020;10(1):14362.
32. Sakamoto A, Kawakami R, Mori M, et al. CD163+ macrophages restrain vascular calcification, promoting the development of high-risk plaque. *JCI Insight*. 2023;8(5):e154922.
33. Wang Y, Liu J, Akatsu C, et al. LAPTM5 mediates immature B cell apoptosis and B cell tolerance by regulating the WWP2-PTEN-AKT pathway. *Proc Natl Acad Sci U S A*. 2022;119(36):e2205629119.
34. Pan J, Zhang M, Dong L, et al. Genome-scale CRISPR screen identifies LAPTM5 driving lenvatinib resistance in hepatocellular carcinoma. *Autophagy*. 2023;19(4):1184-1198.
35. Inoue J, Misawa A, Tanaka Y, et al. Lysosomal-associated protein multispanning transmembrane 5 gene (LAPTM5) is associated with spontaneous regression of neuroblastomas. *PLoS One*. 2009;4(9):e7099.
36. Zhang MM, Liang MJ, Zhang DM, et al. The function and mechanism of LAPTM5 in diseases. *Biomed Pharmacother*. 2024;178:117237.
37. Glowacka WK, Alberts P, Ouchida R, Wang JY, Rotin D. LAPTM5 protein is a positive regulator of proinflammatory signaling pathways in macrophages. *J Biol Chem*. 2012;287(33):27691-27702.
38. Song Z, Wang X, He L, Chen L, Ren Z, Song S. Suppression of lysosomal-associated protein transmembrane 5 ameliorates cardiac function and inflammatory response by inhibiting the nuclear factor-kappa B (NF-κB) pathway after myocardial infarction in mice. *Exp Anim*. 2022;71(4):415-425.
39. Chen X, Zhu S, Huang C, Liu J, Wang J, Cui S. Bioinformatic analyses reveal lysosomal-associated protein transmembrane 5 as a potential therapeutic target in lipotoxicity-induced injury in diabetic kidney disease. *Ren Fail*. 2024;46(2):2359638.
40. Hua W, Ma S, Pang Y, et al. Intracerebral hemorrhage-induced brain injury: the role of lysosomal-associated transmembrane protein 5. *Mol Neurobiol*. 2023;60(12):7060-7079.
41. Zhang Z, Wang L, Wang Z, et al. Lysosomal-associated transmembrane protein 5 deficiency exacerbates cerebral ischemia/reperfusion injury. *Front Mol Neurosci*. 2022;15:971361.
42. Jiang L, Zhao J, Yang Q, et al. Lysosomal-associated protein transmembrane 5 ameliorates non-alcoholic steatohepatitis by promoting the degradation of CDC42 in mice. *Nat Commun*. 2023;14(1):2654.
43. Thompson N, Isenberg DA, Jury EC, Ciurtin C. Exploring BAFF: its expression, receptors and contribution to the immunopathogenesis of Sjögren's syndrome. *Rheumatol Oxf*. 2016;55(9):1548-1555.
44. Marín-Rosales M, Cruz A, Salazar-Camarena DC, et al. High BAFF expression associated with active disease in systemic lupus erythematosus and relationship with rs9514828C > T polymorphism in TNFSF13B gene. *Clin Exp Med*. 2019;19(2):183-190.
45. Wei F, Chang Y, Wei W. The role of BAFF in the progression of rheumatoid arthritis. *Cytokine*. 2015;76(2):537-544.
46. Chen R, Wang X, Dai Z, et al. TNFSF13 is a novel onco-inflammatory marker and correlates with immune infiltration in gliomas. *Front Immunol*. 2021;12:713757.
47. Wang Z, Wang Y, Cui Y, et al. Association of serum BAFF levels with cardiovascular events in ST-segment elevation myocardial infarction. *J Clin Med*. 2023;12(4):1692.
48. Mattiola I, Mantovani A, Locati M. The tetraspan MS4A family in homeostasis, immunity, and disease. *Trends Immunol*. 2021;42(9):764-781.
49. Sanyal R, Polyak MJ, Zuccolo J, et al. MS4A4A: a novel cell surface marker for M2 macrophages and plasma cells. *Immunol Cell Biol*. 2017;95(7):611-619.
50. Mattiola I, Tomay F, De Pizzol M, et al. The macrophage tetraspan MS4A4A enhances dectin-1-dependent NK cell-mediated resistance to metastasis. *Nat Immunol*. 2019;20(8):1012-1022.
51. Li Y, Shen Z, Chai Z, et al. Targeting MS4A4A on tumour-associated macrophages restores CD8+ T-cell-mediated antitumour immunity. *Gut*. 2023;72(12):2307-2320.
52. Sui Y, Zeng W. MS4A4A regulates arginase 1 induction during macrophage polarization and lung inflammation in mice. *Eur J Immunol*. 2020;50(10):1602-1605.
53. Lee JH, Kim JW, Kim DK, et al. The Src family kinase Fgr is critical for activation of mast cells and IgE-mediated anaphylaxis in mice. *J Immunol (Baltimore, Md: 1950)*. 2011;187(4):1807-1815.
54. Acín-Pérez R, Iborra S, Martí-Mateos Y, et al. Fgr kinase is required for proinflammatory macrophage activation during diet-induced obesity. *Nat Metab*. 2020;2(9):974-988.
55. Liu Y, Yang H, Luo N, et al. An Fgr kinase inhibitor attenuates sepsis-associated encephalopathy by ameliorating mitochondrial dysfunction, oxidative stress, and neuroinflammation via the SIRT1/PGC-1α signaling pathway. *J Transl Med*. 2023;21(1):486.
56. Wang A, Kang X, Wang J, Zhang S. IFIH1/IRF1/STAT1 promotes sepsis associated inflammatory lung injury via activating macrophage M1 polarization. *Int Immunopharmacol*. 2023;114:109478.
57. Huang SL, Chang TC, Chao CCK, Sun NK. TLR4/IL-6/IRF1 signaling regulates androgen receptor expression: a potential therapeutic target to overcome taxol resistance in ovarian cancer. *Biochem Pharmacol*. 2021;186:114456.
58. Fan X, Li Q, Wang Y, et al. Non-canonical NF-κB contributes to endothelial pyroptosis and atherogenesis dependent on IRF-1. *Transl Res*. 2023;255:1-13.
59. Cho Y, Kim H, Yook G, et al. Predisposal of interferon regulatory factor 1 deficiency to accumulate DNA damage and promote osteoarthritis development in cartilage. *Arthritis Rheumatol Hoboken*. 2024;76(6):882-893.

How to cite this article: Wang L, Zhang Y, Ji F, et al. Identification of crucial genes for polycystic ovary syndrome and atherosclerosis through comprehensive bioinformatics analysis and machine learning. *Int J Gynecol Obstet*. 2025;170:312-325. doi:10.1002/ijgo.70014