

Identification of Immune-Associated Genes in Diagnosing Polycystic Ovary Syndrome with Metabolic Syndrome by Weighted Gene Co-expression Network Analysis

Yi Xu¹, Yihui Gu², Yang Shen^{1*}, Wen Feng^{2*}

ABSTRACT

Background: Dysregulated immune system and metabolic dysfunction take an essential part on the pathogenesis of polycystic ovary syndrome (PCOS) and metabolic syndrome (MS). The purpose of this study is to identify the key immune-Associated genes of PCOS patients with MS.

Methods: Three PCOS and one MS dataset from the Gene Expression Omnibus (GEO) database were downloaded. After the data sets were combined with the de-batch effect, analysis was conducted using methods such as differentially expressed genes (DEGs), weighted gene co-expression networks (WGCNA), functional enrichment, protein-protein interaction (PPI) network construction, and LASSO regression, which were used to identify candidate immune-related essential genes for diagnosing the combination of these two diseases. Evaluate the obtained genes through ROC and ultimately explore abnormalities in PCOS through immune cell infiltration analysis.

Results: The merged PCOS dataset contains 1513 differentially expressed genes (DEGs), with 691 identified by MS. DEGs of MS were primarily enriched in immune regulation and metabolic dysfunction. The crossed gene functions of DEGs in PCOS and module genes in MS are also mainly enriched in the immune system and metabolic dysfunction. We screened 35 node genes from 57 cross genes through the PPI network and selected six candidate genes using the lasso regression method. Five genes analyzed by the ROC curve had high diagnostic values (RAB4B, SFP91, ARHGAP33, LIME1, and SHMT2) (area under the curve from 0.72 to 0.89). Finally, an imbalance in the proportion of immune cells in PCOS was also observed.

Conclusion: Five candidate genes were identified as the final diagnostic markers. Our conclusion can provide potential diagnostic candidate genes for patients of MS with PCOS.

INTRODUCTION

About 5-18% of women are affected by Polycystic Ovary Syndrome (PCOS) Meier et al. (2018), Wei et al. (2022). The pathogeny is complicated, including genetic and epigenetic susceptibility, hypothalamus and ovarian dysfunction, androgen overexposure, insulin resistance, and obesity-related mechanisms Escobar-Morreale et al. (2018), Qi et al. (2019). Metabolic syndrome (MS) is a chronic non-infectious syndrome, and its clinical features include a series of risk factors, including insulin resistance, elevated blood pressure, abdominal obesity, abnormal Blood glucose, and plasma lipids Bishehsari et al. (2020), Gamble et al. (2014). Therefore, PCOS and MS are closely related but are different entities. In addition, they are both immune-related diseases.

Existing studies have proved that metabolic abnormality correlated factors or their genetic markers are higher in PCOS patients Meier et al. (2018), Stener-Victorin et al. (2021). The incidence rates of glucose intolerance and diabetes Karakas et al. (2010), atherogenic dyslipidemia Bourgneuf et al. (2021), systemic inflammation Qi et al. (2019), non-alcoholic liver disease Kumarendran et al. (2018), and coagulation disorders Targher et al. (2014) are higher in PCOS women. According to previous studies, the related pathological changes of MS are closely related to immune system dysfunction Baecher-Allan et al. (2018) and metabolic abnormality Eckel et al. (2005) related molecules. Therefore, immune system disorders and metabolic abnormality molecules may contribute to the accurate diagnosis of PCOS patients with MS.

¹Department of Obstetrics and Gynaecology, Zhongda Hospital, School of Medicine, Southeast University, Nanjing, China

²Department of Gynecology, The First People's Hospital of Lianyungang, Jiangsu, China.

Correspondence to: Wen Feng, MD, Department of Gynecology, The First People's Hospital of Lianyungang, Lianyungang, 222000, Jiangsu Province, China, China. E-mail: fengw125@126.com

Yang Shen, M.D., Ph.D. Department of Obstetrics and Gynaecology, Zhongda Hospital, School of Medicine, Southeast University, Nanjing 210009, China. E-mail: shenyang@seu.edu.cn

Key Words: PCOS, MS, WGCNA, Immune Infiltration, Diagnosis

There is controversy about the diagnostic criteria of adolescent PCOS patients. Some diagnostic criteria conflict with the normal physiology of the mature reproductive system, and the diagnostic criteria used to diagnose adult women do not apply to adolescents Meier et al. (2018) PCOS is an essential factor leading to women's infertility Goodarzi et al. (2011). There is a close relationship between PCOS and MS He et al. (2019), Sanchez-Garrido et al. (2020), so it is vital to find sensitive and specific diagnostic tools for PCOS early, especially for adolescent patients with MS.

Nevertheless, there is limited research on the diagnostic criteria for MS combined with PCOS to date. Previously, some researchers explored the plasma biomarkers of MS in PCOS patients by metabolomics Zhao et al. (2021). However, the diagnostic genes related to transcriptome have not been explored. Here, we first downloaded the data sets GSE34526, GSE5850, GSE102293(PCOS), and GSE98895(MS) from the Gene Expression Synthesis (GEO) database, identified differentially expressed genes (DEGs) by Limma and found significant module genes by weighted gene co-expression network analysis (WGCNA). Functional enrichment analysis, construction of protein-protein interaction network, The Least Absolute Shrinkage and Selection Operator (LASSO), and analysis of immune cell infiltration are identified the related genes, receiver operating characteristic (ROC) curve is used to evaluate the hub immune-related diagnostic biomarkers of PCOS, and MS. This research may be conducive to determine the immune-related potential diagnostic markers of PCOS in MS patients.

MATERIALS AND METHODS

Microarray Data

Three raw datasets (GSE34526 Kaur et al. (2012) , GSE5850 Wood et al. (2007) , GSE102293 Adams et al.(2006)) of PCOS datasets were selected, including gene expression data, and one dataset(GSE98895 D'Amore et al.(2018)) of MS were downloaded from the GEO database(<https://www.ncbi.nlm.nih.gov/geo/>) Barrett et al. (2013) (Figure 1).

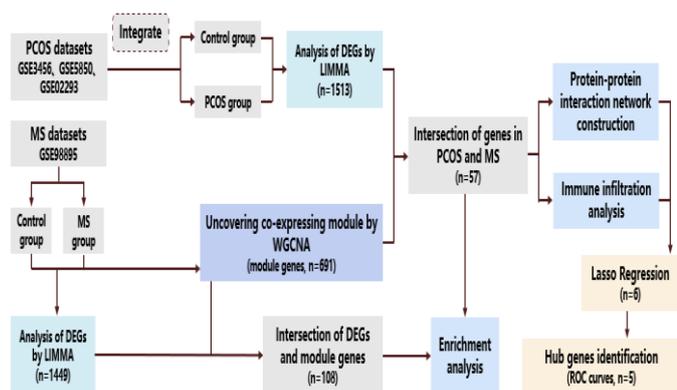


FIGURE 1: Flowchart. GSE, gene expression omnibus series; DEGs, differentially expressed genes; Limma, linear models for microarray data; WGCNA, weighted gene co-expression network analysis.

Merge datasets, control batch and preprocessing

Merge the datasets and control the batch effect Varma et al. (2020). Limma fits a linear equation to the expression of each gene by a generalized linear model. We use the R software package Limma (version 3.40.6) to analyze the differences and get the differential genes between PCOS and the normal group. We set the expression spectrum data set according to $|\log_2 \text{ times change (FC)}| > 1.5$ and $p\text{-value} < 0.05$ as the standard for identifying deg with the Limma package Leek et al. (2012).

Weighted Gene Co-Expression Network Analysis

Weighted gene co-expression network analysis (WGCNA) is suitable for studying biological networks based on the pairwise correlation between variables Langfelder et al. (2008). On the basis of gene correlation, it allows us to define modules (gene clusters) and explore the correlation between co-expressed modules. Also, WGCNA can be used as a feature selection method to analyze the potential relationship between phenotypes and gene modules to help screen out gene networks associated with specific phenotypes.

First, we use the gene expression profile, eliminate the top 50% of genes with the smallest MAD, and remove the outliers and samples. Further, use WGCNA to build a scale-free co-expression network,

Functional Enrichment Analysis

KEGG rest API was used to obtain the latest gene annotation Kanehisa et al. (2017), Using the GO annotations of genes in the R software package (version 3.1.0) as the background and conducting enrichment analysis using the R software package clusterProfiler (version 3.14.3) to obtain the results of gene set enrichment Mayakonda et al. (2018). Set the minimum gene set to 5 and the maximum gene set to 5000, P value of < 0.05 and an FDR of < 0.25 , where considered statistically significant. Interacted analysis was conducted between the DEGs of MS and the most significant module genes of MS's WGCNA, and the DEGs of PCOS have interacted with the most significant mode genes of MS's WGCNA. Finally, the results obtained from the two gene sets were subjected to GO and KEGG analysis.

Protein-Protein Interaction Network Construction

The interaction between proteins is essential at the cellular functional level.

To further explore the protein network of interacted gene expression between PCOS' DEGs and MS' GCNA, the String database was used to establish the protein-protein interaction (PPI) network structure was established Szklarczyk et al. (2021). Select genes with interactions in the PPI network for score sorting, and perform subsequent Lasso regression analysis on genes with values >0.

LASSO regression analysis

Lasso Regression is a variant of linear regression, which constrains the model through specific regularization methods to avoid overfitting and improve generalization ability Yang et al. (2018). The R software package glmnet was used to integrate survival time, survival status, and gene expression data, and regression analysis was performed using the lasso cox regression method. In addition, we also set up a 3-fold cross-validation to obtain the optimal model. The λ value was set at 0.087 ;five genes were ultimately identified.

Receiver Operating Characteristic Evaluation

The R software package proc (version 1.17.0.1) was used for ROC analysis to calculate the area under the curve (AUC value) and its 95% confidence interval. Generally speaking, AUC > 0.7 is considered an excellent diagnostic value Obuchowski et al. (2018).

Immune Infiltration Analysis

CIBERSORT is a cell-type component identification method based on gene expression profile data. It can predict the proportion of various cell subpopulations in the sample, thereby analyzing the number and proportion of different types of cells in a particular organization or sample Newman et al.(2015). The “Cibersort” R software package was used to analyze the proportion of immune cell infiltration between PCOS patients and control groups in order to compare the differences in the proportion of various types of immune cells between PCOS and control groups. The comparison of the proportion of different types of immune cells was visualized by Vioplot Hu et al. (2020).

Statistical Analysis

In this research, all data analysis was conducted using R software. A value with $P < 0.05$ is considered significant. ROC analysis was used to obtain the AUC value

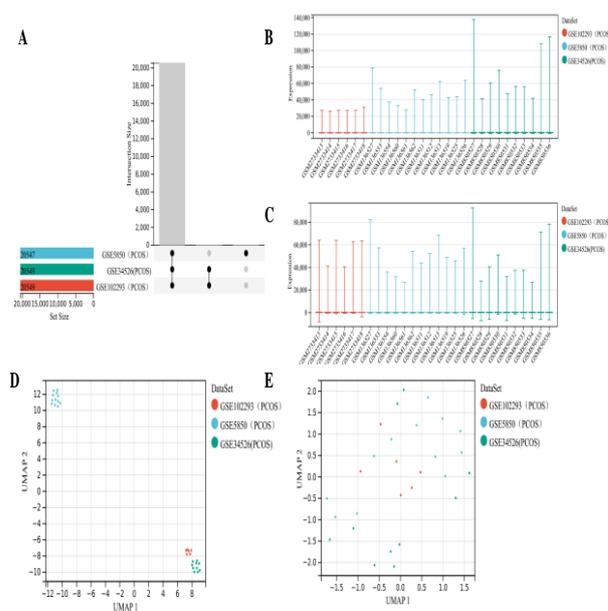
RESULTS

Intersection between data sets and control batch

Figure 2A shows the intersection between 3 PCOS data sets (GSE34526, GSE5850, GSE102293).

Boxplots of gene expressions before and after standardization for 3 data sets, respectively (Figure 2B,2C). UMAP of gene expressions before and after standardization for 3 data sets, respectively (Figure 2D,2E). After controlling the batch effect, the data distribution between each data set tends to be consistent.

Figure 2: Before and after Controlling batch effects on PCOS datasets



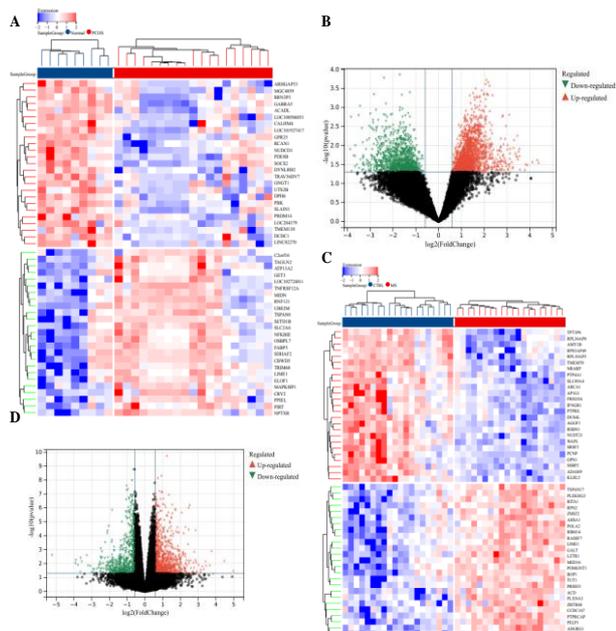
(A) Combined histogram of three data sets for PCOS. (B) Boxplot before batch control. (C) Boxplot after batch control. (D) UMAP before batch control. (E) UMAP after batch control.

Identifying differential genes

Using the Limma package, 1513 DEGs were identified in PCOS datasets, of which 781 were up-regulated, and 732 were down-regulated. The thermal map and volcanic map of PCOS DEGs are shown in Figures 3 A, B. Regarding the MS data set, 1449 DEGs (844 up-regulated and 605 down-regulated) were screened out (Figure 3 C, D).

Figure 3: Differentially expressed genes identification from the integrated PCOS dataset and MS dataset.

(A) Heatmap of the integrate three datasets of GSE34526, GSE5850, GSE102293. (B) Volcano map of the integrate three datasets of GSE34526, GSE5850, GSE102293. (C) Heatmap of GSE98895. (D) Volcano map of GSE98895.



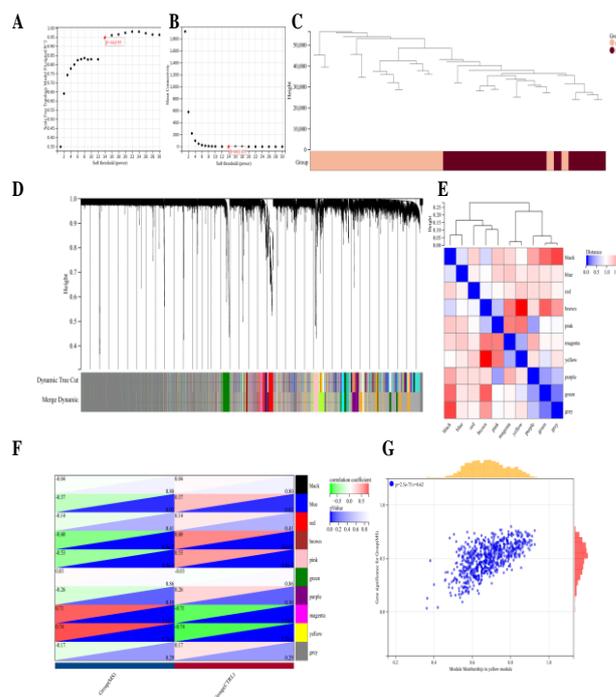
Weighted Gene Co-Expression Network Analysis

WGCNA is used to identify the most significantly correlated module in MS, displayed as a yellow module. We chose $b=14$ (scaleless $R^2=0.9$) as the 'soft' threshold.

(Figures 4A, B). The clustering dendrogram of the MS group and the control group are shown in Figure 4C, which further generated ten gene co-expression modules, each with different colors (Figure 4D, E). The correlation between MS and module genes is shown in Figure 4F, with the yellow module (691 genes) containing the highest coefficient (correlation coefficient=0.74, $p=5.2 \times 10^{-8}$). Therefore, further analysis will be conducted on the yellow module in the future. The members and genes in the MS yellow module have the highest significance ($r=0.62$) (Figure 4G). Therefore, the correlation between this module gene and MS is most significant. Therefore, the correlation between this module gene and MS is most significant.

Figure 4: Module gene of MS obtained by WGCNA method.

(A, B) Analysis of combining scale independence with average connectivity. (C) Clustering dendrogram between MS and control groups. (D) Co-expression module of gene tree. (E) Heatmap of eigengene adjacency. (F) Heatmap of the association between modules and MS. The yellow module is shown to be significantly related to MS. (G) Correlation diagram between module members and gene significance in yellow module.

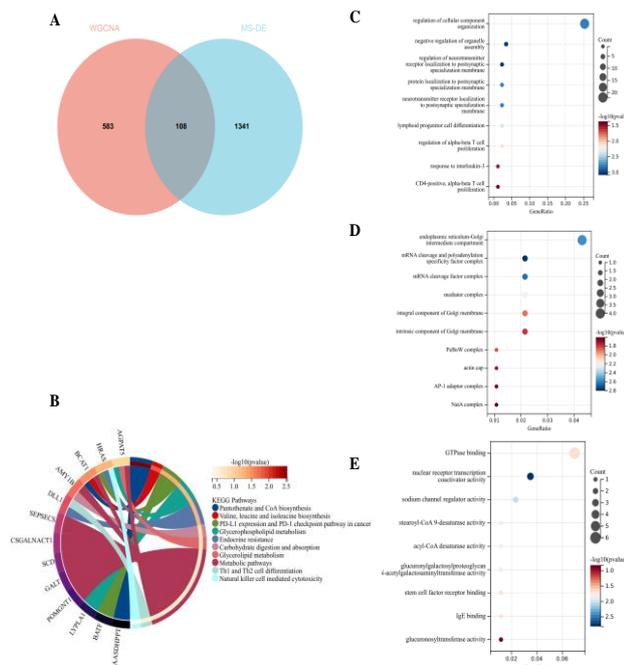


Functional Enrichment Analysis of MS

In order to further explore whether there is a certain correlation between the pathogenesis of MS and PCOS, we conducted an interactive analysis between the DEGs genes of MS and the critical module genes of WGCNA, followed by functional enrichment of the interacted genes. By crossing 1449 DEG and 691 genes in the yellow module, a total of 108 common genes were screened out. Finally, 108 interacting genes were screened out (Figure 5A).

CGs were primarily enriched in the “PD-L1 expression and PD-1 checkpoint pathway in cancer,” “Fatty acid metabolism,” “Th1 and Th2 cell differentiation,” “Metabolic pathways,” and “Endocrine resistance” from KEGG analysis (Figure 5B). CGs were mainly enriched in biological process (BP) terms, including “protein localization to postsynaptic specialization membrane,” “regulation of alpha-beta T cell proliferation,” “response to interleukin-3,” and “CD4-positive, alpha-beta T cell proliferation” from GO analysis (Figure 5C). the CGs were mainly located in the “mRNA cleavage and polyadenylation specificity factor complex,” “mediator complex”, and “integral component of Golgi membrane” from cellular component (CC) ontology (Figure 5D). “GTPase activity,” “carbohydrate derivative binding,” “immunoglobulin binding,” and “immunoglobulin binding” were the most significant items in CGs from Molecular function (MF) analysis (Figure 5E). CGs of MS were mainly related to immunity and metabolic reaction.

Figure 5: Enrichment and WGCNA analysis.



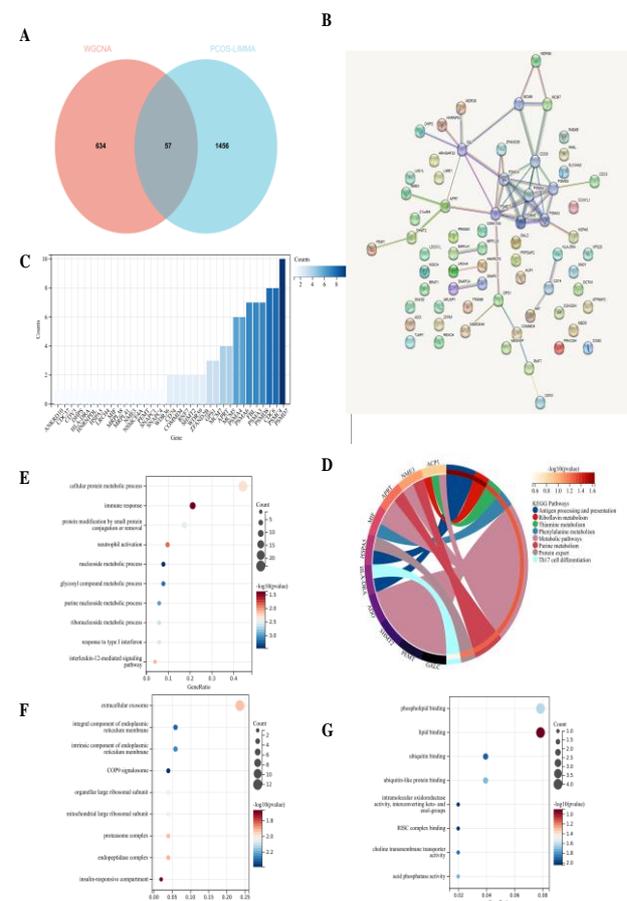
(A) 108 genes interacted by MS DEGs and WGCNA yellow module genes of venn diagram. (B) KEGG pathway analysis of the intersected genes. (C–E) GO analysis of the intersected genes, including biological process (BP), cellular component (CC), and molecular function (MF), respectively from GO analysis.

Functional Enrichment Analysis of PCOS With MS and Identification of Candidate Genes

In order to further explore the association between MS essential genes and the pathogenesis of PCOS, 57 genes were ultimately selected by interacting with the DEGs of PCOS and the critical module genes of MS (Figure 6A). Fifty-seven genes were primarily enriched in the “Metabolic pathways,” “Antigen processing and presentation,” and “Allograft rejection”; all of the above ontologies were intimately related to the immune function and metabolic pathway from KEGG enrichment analysis (Figure 6D). Also, genes were enriched in “Allograft rejection,” “glycosyl compound metabolic process,” and “interleukin-12-mediated signaling pathway” (BP); “extracellular exosome,” “proteasome complex,” and “insulin-responsive compartment” (CC); and “lipid binding,” “intramolecular oxidoreductase activity, interconverting keto- and enol-groups,” and “acid phosphatase activity” (MF) from GO analysis (Figures 6E–G).

The enriched pathway of interacting genes is closely related to immunity, and then node genes that interact with the genes are mined through PPI protein networks for subsequent filtering analysis. Figure 6B shows a close interaction between 35 genes; The genes are sorted according to the counts in Figure 6C.

FIGURE 6: Enrichment analysis of intersected genes between WGCNA yellow module gene and DEGs of PCOS.



(A) Venn diagram shows that 57 genes were intersected by DEGs of PCOS and yellow module gene of WGCNA. (B) PPI network reveals the interaction of 35 genes. (C)

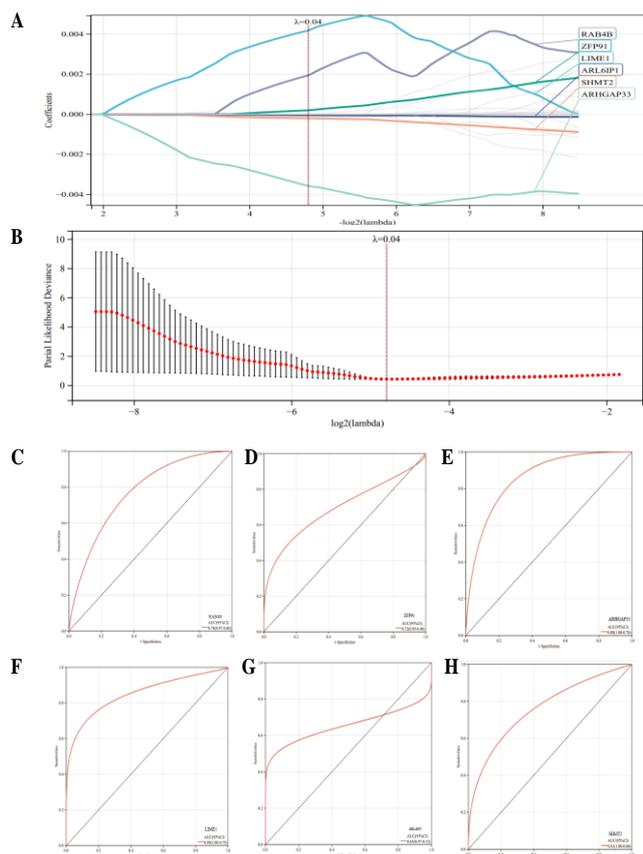
The bar chart shows the gene nodes of 35 genes in PPI network. (D) KEGG analysis of 57 intersected genes. (E–G) GO analysis of 57 intersected genes.

LASSO regression and ROC analysis

Figure 7A and B show that the LASSO regression algorithm identified six biomarkers with diagnostic values (RAB4B, SFP91, ARHGAP33, LIME1, ARL6PI, and SHMT2). Subsequent ROC analysis was performed, and the ARL6PI gene (AUC=0.65) was removed. The AUC values of the other five genes were all >0.7 (Figure 7C–H).

FIGURE 7: Lasso regression analysis.

(A, B) Biomarkers screening in the Lasso model, six diagnostic genes suitable for PCOS patients with MS were screened out. (C–H) The ROC curve of each candidate gene (RAB4B, SFP91, ARHGAP33, LIME1, ARL6PI, and SHMT2) show the diagnostic value, the AUC value of five genes is > 0.7, except ARL6PI.



Immune Cell Infiltration Analysis

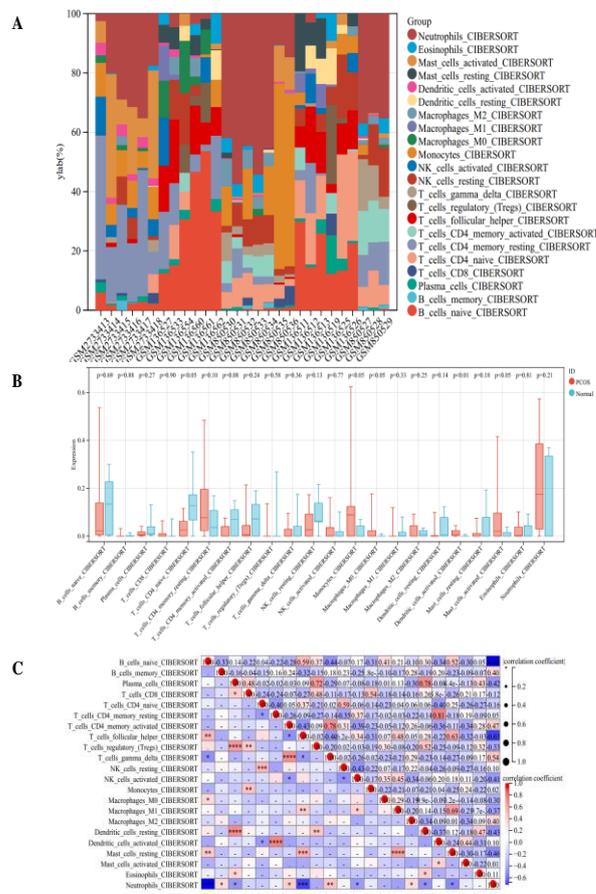
From the above conclusion, we observe that the enriched functions of MS-related genes are closely related to the pathogenesis of PCOS. Both diseases are highly enriched in immune function, so that we will conduct immune-related analysis on PCOS below. The proportion of 22 kinds of immune cells in the PCOS group and control group samples is different, which is shown in the bar plot (Figure 8A). The boxplot demonstrated that PCOS patients had a higher level of monocytes, M0 macrophages, activated dendritic cells, and activated mast cells, and a lower level of native CD4 T cells (Figure 8B).

Correlation analysis between immune cells revealed that resting memory CD4 T cells were positively associated with activated dendritic cells ($r = 0.81$). That activated memory CD4 T cells were positively related to $\delta\gamma$ T cells ($r = 0.78$), whereas native B cells were negatively related to neutrophils ($r = -0.68$) (Figure 8B). In summary, the proportion of immune cell infiltration in patients with polycystic ovary syndrome significantly changes, and the correlation between these changes is significant, which may be a potential regulatory point for the treatment of polycystic ovary syndrome.

FIGURE 8: Immune cell infiltration analysis of PCOS.

The barplot shows the proportion of 22 kinds of immune cells in the merged PCOS datasets.

(B) Immune checkpoints between PCOS and control groups visualized by the vioplo diagram. (C) Correlation between 22 kinds of immune cells.



DISCUSSION

PCOS is one of the leading causes of infertility in women. There is no previous study to find diagnostic genes using the transcriptome data of two diseases. Recent studies have identified several biomarkers for PCOS diagnosis, such as sCD36 Glintborg et al.(2008),SHBG Qu et al.(2020), and some adipocytokines(adiponectin, visfatin, and vaspin and apelin), copeptin, irisin, PAI-1 and zonulin Polak et al.(2017). In this study, we used a series of integrated bioinformatics analyses to identify candidate biomarkers for PCOS in MS patients. Our final result was the identification of five essential immune-related candidate genes (RAB4B, SFP91, ARHGAP33, LIME1, and SHMT2).

We first selected candidate genes through methods such as differential genes and WGCNA to demonstrate a close correlation between the pathogenesis of the two diseases. Then, we used LASSO regression and ROC analysis to ultimately evaluate the accuracy of the candidate genes.

When the AUC value of the ROC curve > 0.7 , it implies accuracy Heagerty et al.(2005). When the gene expression is high, we can determine whether PCOS is present based on the gene expression results of MS patients.

Ras-Related Protein Rab-4B (RAB4B) is a member of the RAS superfamily of small GTPases involved in vesicular trafficking He et al.(2002). Its functions mainly focus on the innate immune system and protein metabolism. RAB4B is a small GTP enzyme that primarily maintains the localization of glucose transporter GLUT4 in adipocytes in terms of function. The RAB4B dysfunction of T cells will lead to an imbalance of Treg/Th17 in fat, leading to pathological changes such as adipose tissue dysfunction and insulin resistance Gilleron et al.(2018). Large amounts of studies show that strong correlation between insulin resistance, lipid metabolic dysfunction, and polycystic ovarian syndrome (PCOS) Escobar-Morreale et al. (2018), Gilleron et al.(2018). RAB4B is also overexpressed in PCOS. Thus we suggest that RAB4B could represent a potential diagnostic target for PCOS in MS patients.

ZFP91 can cause abnormal expression of FOXA1, leading to increased proliferation and migration activity and increased chemotherapy resistance Wang et al.(2021). ZFP91 is related to drug resistance of T-cell lymphoma Wu et al.(2022). Thus, ZFP91 significantly regulates the homeostasis and function of T cells Wang et al.(2021). Peripheral blood Tregs in patients with PCOS decreased compared to the normal group Krishna et al.(2015). ZFP91 take a significant impact on tumorigenesis and NF- κ B pathways Hanafi et al.(2021). Many signal pathways are affected by OS, such as Keap1-Nrf2, NF- κ B, FOXO, and MAPK pathways Hanafi et al.(2021). Oxidative stress (OS) is closely related to PCOS Li et al.(2022). ZFP91 is overexpressed in MS patients with PCOS. Therefore, we speculate that ZFP91 may induce immune function changes and pathological processes such as OS, leading to PCOS. In addition, we infer that ZFP91 can be used as a biomarker for diagnosis.

The pathways involved in the ARHGAP33 (Rho GTPase Activating Protein 33) gene include Rho GTPase signaling and RAC1 GTPase cycling. Chiang et al.(2003). It may be involved in the process of intracellular transportation Nakazawa et al.(2016), and participate in the process of insulin-regulating glucose transport Chiang et al.(2003). Previously, a large number of experimental and clinical data showed that compared with healthy controls, PCOS patients had abnormal endometrial expression disorder, increased endometrial insulin resistance, impaired glucose transport and utilization, followed by chronic inflammation, immune dysfunction, changes in uterine blood vessel distribution and other abnormal changes

Meier et al.(2018).Therefore, we speculate that ARHGAP33 can be used as a biomarker for diagnosis.

LIME1 encodes a transmembrane adapter protein that can bind to the Src family kinases Lck and Lyn, thereby connecting downstream signaling pathways after T and B cell signal stimulation Park et al.(2020), Ahn et al.(2006). It participates in signal transduction mediated by BCR and TCR. In the absence of TCR signaling, it may be involved in CD4-mediated inhibition of T-cell activation. T-cell dysfunction may play a role in the pathogenesis of PCOS. Therefore, LIME1 is obviously related to PCOS.

Previous studies have demonstrated that immune dysfunction and metabolic disorders are closely related to the pathological mechanism of PCOS Escobar-Morreale et al. (2018), Meier et al.(2018). An abundant number of studies have demonstrated that insulin resistance and immune system dysfunction play an important role in PCOS. According to our study, the pathway functions enriched by DEGs in PCOS patients are mainly in immune function and metabolism. PCOS patients had a higher level of monocytes, M0 macrophages, activated dendritic cells, and activated mast cells, and a lower level of native CD4 T cells, which is consistent with previous research results. Above all, identifying biomarkers can provide precise treatment for the diagnosis and targeted treatment of PCOS.

DECLARATIONS

Conflict of Interest

There are no conflicts of interest regarding the publication of this article.

Authors' contributions

Yi Xu conceived and designed the study, drafted the article, revised the article critically, and be responsible for materials.

Yihui Gu had final approval of the submitted versions.

Funding

This study was supported by National Natural Science Foundation of China (No. 82072078), Jiangsu Province Key Research and Development Project (SBE2020741118), and Postgraduate Research & Practice Innovation Program of Jiangsu Province (SJCX22_0070).

Ethics approval and consent to participate

Not applicable.

Availability of data and materials

The following information was supplied regarding data availability: Data is available at the GEO database (<https://www.ncbi.nlm.nih.gov/geo/>).

Acknowledgements

Not applicable

Consent for publication

Not applicable

REFERENCES

1.Meier R K. 2018 Sep. Polycystic ovary syndrome. *Nurs Clin North Am.* 53(3):407-420 .

2.Weï H, Huo P, Liu S, et al. 2022 Oct 7. Posttranslational modifications in pathogenesis of PCOS. *Front Endocrinol (Lausanne).* 13:1024320.

3.Escobar-Morreale H.F. 2018 May. Polycystic ovary syndrome: definition, aetiology, diagnosis and treatment. *Nat Rev Endocrinol.*14(5):270-284.

4.Qi X, Yun C, Sun L, et al. 2019 Aug. Gut microbiota-bile acid-interleukin-22 axis orchestrates polycystic ovary syndrome. *Nat Med.* 25(8):1225-1233.

5.Bishehsari F,Voigt RM, Keshavarzian A, et al. 2020 Dec. Circadian rhythms and the gut microbiota: from the metabolic syndrome to cancer. *Nat Rev Endocrinol.* 16(12):731-739.

6.Gamble K.L ,Berry R ,Frank S J, et al. 2014 Aug. Circadian clock control of endocrine factors. *Nat Rev Endocrinol.* 10(8):466-75.

7.Meier R K. 2018 Sep. Polycystic ovary syndrome. *Nurs Clin North Am.* 53(3):407-420 .

8.Stener-Victorin E, Q. Deng. 2021 Sep. Epigenetic inheritance of polycystic ovary syndrome - challenges and opportunities for treatment. *Nat Rev Endocrinol.* 17(9):521-533.

9.Karakas S.E, Kim K, Duleba A.J, et al. 2010 Apr.Determinants of impaired fasting glucose versus glucose intolerance in polycystic ovary syndrome. *Diabetes Care.* 33(4):887-93.

10.Bourgneuf C, Bailbe D, Lamaziere A, et al. 2021 Feb 16. The Goto-Kakizaki rat is a spontaneous prototypical rodent model of polycystic ovary syndrome. *Nat Commun.* 12(1):1064.

11.Kumarendran B, Rielly M W O, Manolopoulos K N, et al. 2018 Mar 28. Polycystic ovary syndrome, androgen excess, and the risk of nonalcoholic fatty liver disease in women: A longitudinal study based on a United Kingdom primary care database. *PLoS Med.* 15(3):e1002542.

12.Targher G,Zoppini G,Bonora E, et al. 2014 Jul.Hemostatic and fibrinolytic abnormalities in polycystic ovary syndrome. *Semin Thromb Hemost.* 40(5):600-18.

13.Baecher-Allan C, Kaskow B.J, Weiner H.L, et al. 2018 Feb 21. Multiple Sclerosis: Mechanisms and Immunotherapy. *Neuron.* 97(4):742-768.

14.Eckel R.H, Grundy S.M, Zimmet P. Z, et al. 2005 Apr. The metabolic syndrome. *Lancet.* 365(9468):1415-28.

15.Goodarzi M.O, Dumesic D A, Chazenbalk G, et al. 2011 Apr. Polycystic ovary syndrome: etiology, pathogenesis and diagnosis. *Nat Rev Endocrinol.* 7(4):219-31.

16.He Y, Lu Y, Zhu Q, et al. 2019 Aug. Influence of metabolic syndrome on female fertility and in vitro fertilization outcomes in PCOS women. *Am J Obstet Gynecol.* 221(2):138.e1-138.e12.

17.Sanchez-Garrido M.A, Tena-Sempere M. 2020 May.Metabolic dysfunction in polycystic ovary syndrome: Pathogenic role of androgen excess and potential therapeutic strategies. *Mol Metab.* 35:100937.

18.Zhao X, Feng X, Zhao X, et al. 2021 Jun 2 How to Screen and Prevent Metabolic Syndrome in Patients of PCOS Early: Implications From Metabolomics. *Front Endocrinol (Lausanne).* 12:659268.

19.Kaur S, Archer K J, Devi M G, et al. 2012 Oct .Differential gene expression in granulosa cells from polycystic ovary syndrome patients with and without insulin resistance: identification of susceptibility gene sets through network analysis. *J Clin Endocrinol Metab.* 97(10):E2016-21.

20.Wood J.R, Dumesic D.A, Abbott D.H, et al. 2007 Feb. Molecular abnormalities in oocytes from women with polycystic ovary syndrome revealed by microarray analysis. *J Clin Endocrinol Metab.* 92(2):705-13.

21.Adams J,Liu Z, Ren Y.A, et al. 2016 Sep. Enhanced Inflammatory Transcriptome in the Granulosa Cells of Women With Polycystic Ovarian Syndrome. *J Clin Endocrinol Metab.* 101(9):3459-68.

22.D'Amore S, Hardfeldt J, Cariello M, et al. 2018 Jul 1. Identification of miR-9-5p as direct regulator of ABCA1 and HDL-driven reverse cholesterol transport in circulating CD14+ cells of patients with metabolic syndrome. *Cardiovasc Res.* 114(8):1154-1164.

23.Barrett T, Wilhite S.E, Ledoux P, et al. 2013 Jan. NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res.* 41(Database issue):D991-5.

24.Varma S. 2020 Apr 9. Blind estimation and correction of microarray batch effect. *PLoS One.* 15(4):e0231446.

25. Leek J.T, Johnson W.E, Parker H.S, et al. 2012 Mar 15. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*. 28(6):882-3.
26. Langfelder P, Horvath S. 2008 Dec 29. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 9:559.
27. Kanehisa M, Furumichi M, Tanabe M, et al. 2017 Jan 4. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res*. 45(D1):D353-D361.
28. 2017 Jan 4. Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res*. 45(D1):D331-D338.
29. Mayakonda A, Lin D.C, Assenov Y, et al. 2018 Nov. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res*. 28(11):1747-1756.
30. Szklarczyk D, Gable A.L, Nastou K.C, et al. 2021 Jan 8. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res*. 49(D1):D605-D612.
31. Yang C, Delcher C, Shenkman E, et al. 2018 Nov 20. Machine learning approaches for predicting high-cost high need patient expenditures in health care. *Biomed Eng Online*. 17(Suppl 1):131.
32. Obuchowski N.A, Bullen J.A. 2018 Mar 29. Receiver operating characteristic (ROC) curves: review of methods with applications in diagnostic medicine. *Phys Med Biol*. 63(7):07TR01.
33. Newman A.M, Liu C.L, Green M.R, et al. 2015 May. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 12(5):453-7.
34. Hu K. 2020 Sep 23. Become Competent within One Day in Generating Boxplots and Violin Plots for a Novice without Prior R Experience. *Methods Protoc*. 3(4):64.
35. Glinborg D, Hojlund K, Andersen M, et al. 2008 Feb. Soluble CD36 and risk markers of insulin resistance and atherosclerosis are elevated in polycystic ovary syndrome and significantly reduced during pioglitazone treatment. *Diabetes Care*. 31(2):328-34.
36. Qu X, Donnelly R. 2020 Nov 1. Sex Hormone-Binding Globulin (SHBG) as an Early Biomarker and Therapeutic Target in Polycystic Ovary Syndrome. *Int J Mol Sci*. 21(21):8191.
37. Polak K, Czyzyk A, Simoncini T, et al. 2017 Jan. New markers of insulin resistance in polycystic ovary syndrome. *J Endocrinol Invest*. 40(1):1-8.
38. Heagerty P.J, Zheng Y. 2005 Mar. Survival model predictive accuracy and ROC curves. *Biometrics*. 61(1):92-105.
39. He H, Dai F, Yu Long, et al. 2002. Identification and characterization of nine novel human small GTPases showing variable expressions in liver cancer tissues. *Gene Expr*. 10(5-6):231-42.
40. Gilleron J, Bouget G, Ivanov S, et al. 2018 Dec 18. Rab4b Deficiency in T Cells Promotes Adipose Treg/Th17 Imbalance, Adipose Tissue Dysfunction, and Insulin Resistance. *Cell Rep*. 25(12):3329-3341.e5.
41. Wang A, Ding L, Wu Z, et al. 2021 Feb 1. ZFP91 is required for the maintenance of regulatory T cell homeostasis and function. *J Exp Med*. 218(2):e20201217.
42. Wu W, Nelson G.M, Koch R, et al. 2022 Mar 31. Overcoming IMiD resistance in T-cell lymphomas through potent degradation of ZFP91 and IKZF1. *Blood*. 139(13):2024-2037.
43. Krishna M.B, Joseph A, Subramaniam A.G, et al. 2015 Jan. Reduced Tregs in peripheral blood of PCOS patients - a consequence of aberrant Il2 signaling. *J Clin Endocrinol Metab*. 100(1):282-92.
44. Hanafi M, Chen X, Neamati N, et al. 2021 Feb 11. Discovery of a Napabucasin PROTAC as an Effective Degradator of the E3 Ligase ZFP91. *J Med Chem*. 64(3):1626-1648.
45. Li R, Li Z, Huang Y, et al. 2022 Aug 5. The effect of magnesium alone or its combination with other supplements on the markers of inflammation, OS and metabolism in women with polycystic ovarian syndrome (PCOS): A systematic review. *Front Endocrinol (Lausanne)*. 13:974042.
46. Chiang S.H, Hwang J, Legendre M, et al. 2003 Jun 2. TCGAP, a multidomain Rho GTPase-activating protein involved in insulin-stimulated glucose transport. *Embo j*. 22(11):2679-91.
47. Nakazawa T, Hashimoto R, Sakoori K, et al. 2016 Feb 3. Emerging roles of ARHGAP33 in intracellular trafficking of TrkB and pathophysiology of neuropsychiatric disorders. *Nat Commun*. 7:10594.
48. Park I, Son M, Ahn E, et al. 2020 Nov 30. The Transmembrane Adaptor Protein LIME Is Essential for Chemokine-Mediated Migration of Effector T Cells to Inflammatory Sites. *Mol Cells*. 43(11):921-934.
49. Ahn E, Lee H, Yun Y, et al. 2006 Feb 15. LIME acts as a transmembrane adapter mediating BCR-dependent B-cell activation. *Blood*. 107(4):1521-7.