1. Introduction

Polycystic ovarian syndrome (PCOS) is a relatively common endocrine and metabolic disorder in females, characterized by a complex etiology. It represents a major cause of infertility among women of reproductive age. The main clinical manifestations of polycystic ovary syndrome (PCOS) are irregular menstruation, polycystic ovarian morphology (PCOM), ovulation disorder (OD) and infertility. Most patients with polycystic ovarian syndrome (PCOS) exhibit a range of metabolic abnormalities. These include elevated or excessive body mass index (BMI), acanthosis nigricans, insulin resistance (IR), and altered glucose and lipid metabolism [1]. Among these, IR is particularly significant, with its prevalence ranging from 10-25% in the general population. However, in PCOS patients, this prevalence escalates to 60-70% [2]. IR is a critical factor contributing to the development of various conditions, such as impaired glucose tolerance (IGT), diabetes mellitus (DM), hypertension, hyperlipidemia, abdominal obesity, and cardiovascular and cerebrovascular diseases. Additionally, PCOS increases the risk of gestational diabetes mellitus and gestational hypertension, leading to adverse pregnancy outcomes [3].

The specific etiology of polycystic ovarian syndrome (PCOS) remains unclear, though current understanding implicates a combination of environmental and genetic factors. Environmental influences, particularly the effects of hyperandrogenism during embryonic development, are considered to play a role. Additionally, lifestyle factors such as poor diet have been associated with an increased risk of PCOS [4]. From a genetic perspective, the familial clustering observed in PCOS cases suggests a genetic
Predisposition. Cytogenetic studies indicate that the inheritance of PCOS could follow patterns of X-linked recessive, autosomal dominant, or polygenic transmission [5]. Genome-Wide Association Studies (GWAS) have identified numerous genes associated with PCOS. These genes are involved in a variety of functions, including androgen synthesis and regulation, insulin synthesis, gonadotropin function and regulation, adipose tissue processes, chronic inflammation, and metabolism [3]. Moreover, there are several genes linked to PCOS whose functions remain to be elucidated.

The expansive application of bioinformatics technology in recent years has opened new avenues for the treatment of various diseases. Public databases, such as the Gene Expression Omnibus (GEO), offer a rich resource for research into both oncological and non-oncological diseases. Recent studies focusing on the genetic factors of polycystic ovarian syndrome (PCOS) suggest that gene therapy may emerge as a key area of research [6]. Although some progress has been made in understanding the etiology of PCOS, its specific pathogenesis remains elusive. In this study, we initially screened for differentially expressed genes related to PCOS using the microarray data from GSE5090. Subsequently, we conducted functional enrichment analysis using bioinformatics software. Key genes were then identified through the use of BioGPS, the String database, and protein-protein interaction (PPI) network analysis. The validation of these key genes provides a foundation for developing new insights into the underlying mechanisms of PCOS and formulating novel approaches for its therapeutic intervention.

2. Materials and Methods

2.1. Data sources
In this study, we focused on key genes associated with polycystic ovarian syndrome (PCOS). Data were extracted from the Gene Expression Omnibus (GEO) database, accessible at https://www.ncbi.nlm.nih.gov/geo/, with the data retrieval conducted in August 2023. The dataset utilized, GSE5090, comprises gene chip data pertinent to PCOS, including samples from 7 PCOS patients and 8 healthy controls. We also have included two replicates, one biological and the other technical. This dataset is based on the GPL96 platform. For specific details regarding the dataset, refer to the GEO website (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=gse5090).

2.2. Data processing
Raw data were processed and analysed using R (version 4.3.1). The median value of each sample was normalised for background correction using the between-array limma package. A robust multi-chip average (RMA) was then created and perfect matches of the raw data were log-transformed. The PCOS gene chip data (GSE5090) from the GEO database was analysed to screen for differentially expressed genes. The dataset included 9 PCOS samples and 8 control samples for differential analysis. The criteria for differential gene screening were: $P < 0.05$, $|\text{LogFC}| > 2$.

2.3. Tissue-specific gene expression analysis
Information about gene function can be obtained from relevant tissue-specific genes. Tissue-specific DEGs were screened using the BioGPS database (http://biogps.org/#goto=welcome). Highly tissue-specific transcripts were mapped to individual tissues if all of the following criteria were met: (1) the median expression was 40-fold higher than the median expression in all other tissues; (2) the highest expression level was at least 2-fold higher than the second highest expression level at least 2-fold higher.

2.4. Functional annotation and KEGG pathway analysis
Metascape (http://metascape.org/) is a powerful online software for gene function enrichment analysis, and the tool integrates authoritative databases such as GO, KEGG, and Hallmark. In this study, we used Metascape software to analyse the functional enrichment of differentially expressed genes, and the clustering result of $P < 0.05$ was defined as significant clustering.

2.5. Identification of key genes
In this study, we subjected the screened target genes to PPI network analysis using the STRING (https://string-db.org/) online database. A combined score $> 0.4$ was defined as the threshold. Based on the above study, the CytoHubba plugin in Cytoscape 3.10.0 software was utilised in this study to perform key gene identification, with node degree $\geq 10$ defined as the threshold and a Wayne diagram were plotted, which was used to identify tissue differentially expressed genes versus key genes in the PPI network.

2.6. Column diagram model construction
We constructed a column chart model to predict PCOS risk using the "rms" package. The performance of the column chart model was assessed by calculating the Harrell consistency index, which evaluates the predictive power. We then used the "ROC" package to construct subject operator characteristic (ROC) curves to validate the diagnostic validity of the candidate biomarkers. We used the area under the ROC curve (AUC) to indicate accuracy.

2.7. Statistical analysis
The dataset used contains data from the GEO database (https://www.ncbi.nlm.nih.gov/geo/) downloaded in MINI ML format. mRNA differential expression was investigated using the limma package of R software (version: 3.40.2). Adjusted $P$ values were analysed in GEO to correct for false positive results. "Adjusted $P < 0.05$ with $\log2$ (fold change) $> 1$ or $\log2$ (fold change) $< -1$" was defined as a threshold mRNA differential expression screen. Functional enrichment: To further confirm the potential function of potential targets, data were analysed by functional enrichment. Gene ontology (GO) is a widely used tool for annotating genes with functions, especially molecular functions (MF), biological pathways (BP), and cellular components (CC). KEGG enrichment analysis is practical and can be used to analyse gene functions as well as related high-level genomic functional information. To better understand the role of target genes, the ClusterProfiler package (version: 3.18.0) in R was used to analyse the GO function of potential mRNAs and enrich KEGG pathways. Expression heatmaps were demonstrated by the R package pheatmap.

3. Results

3.1. Differential expression analysis
To explore potential therapeutic targets for PCOS, a bioinformatics approach was used to identify differential-
Therapeutic targets in polycystic ovary syndrome genes

3.2. Specificity of DEGs in tissue expression

Using BioGPS, we screened 88 DEGs that were preferentially expressed in specific tissues. Tissue-specific expression of markers of the immune system (18, 20.5%) varied the most, followed by the digestive system (17, 19.3%), the nervous system (14, 15.9%), the respiratory system (10, 11.3%), and the skin/skeletal muscle system (8, 9.1%), urinary system (6, 6.8%), reproductive system (7, 7.9%), circulatory system (5, 5.7%), and motor system (3, 3.4%) (Table 1).

3.3. Enrichment analysis of DEGs

GO analysis was performed using metascape. The enriched GO terms were classified into 3 categories: BP, CC, and MF. As shown in Table 2 and Figure 3, in the BP group, DEGs were mainly enriched in SRP-dependent co-translated proteins targeting membranes, and co-translated proteins targeting membranes. CC analysis showed that the differentially expressed genes were mainly enriched in "cytoplasmic ribosomes" and "ribosomal subunits". In terms of MF, DEGs most occurred in the composition of the ribosome and antioxidant activity. Using KEGG pathway enrichment analysis of differentially expressed genes. KEGG analysis shows that DEGs mainly occurred in glycine, leucine and isoleucine degradation "frequently" "mediated the ubiquitin-proteasome proteolytic" and "thermogenesis" (Figure 4).

3.4. Construction of PPI network and identification of key genes

The PPI network was constructed using STRING database and Cytoscape v3.10.0. The top 20 hub genes were identified using the Cytohubba plugin according to the proximity algorithm. Extract the hub genes including CD74, RPL30, RPS27, RPS29, RPL41, NBEAL1, RPS24, RPL37, RPS4X, UBB, RPL37, UBC, RPS4X, EEF1A1, SELENOP, PL41 and ANXA1 (Figure 5). Next, we built a Venn diagram to confirm the hub genes. Next, we performed tissue-specific gene expression analysis and enrichment analysis, and constructed a PPI network. In this study, omental adipose tissue gene chip data (GSE5090) of PCOS was collected from the GEO database of NCBI. The data consisted of nine PCOS samples and eight samples from healthy controls. The microarray data from the GSE5090 dataset was normalised (Figure 1). A total of 1324 DEGs were identified after setting cutoff at FDR<0.05 and |log2 (FC)|>1 according to the screening conditions (Figure 2). Seventeen genes significantly expressed between the two groups were extracted using cutoff at |log2 (FC)| >3 and FDR<0.005 (Figure 2A).

Table 1. Tissue-specific genes identified by BioGPS.

<table>
<thead>
<tr>
<th>System</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immune System</td>
<td>RPL30, RPS27, RPL34, LDHB, PARK7, RPL41, NBEAL1, RPS24, RPL37, RPS4X, HMGBI1, CD74, RPS17, RPS15, PRKAR2B, RPL29, RPL39, PPIA</td>
</tr>
<tr>
<td>Respiratory System</td>
<td>S100A6, GPX1, MYL12B, ANXA2, CAV1, NQO1, CAPNS1, PSAP, MT2A, ALDH1A3</td>
</tr>
<tr>
<td>Digestive</td>
<td>ANXA1, IER3, ALDH2, LGALS1, MXRA7, RPS29, CD81, SELENOP, LGALS3, B2M, CES1, PMP22, EEF1A1, CXCL8, IL6, CXCL2, CCL2</td>
</tr>
<tr>
<td>Reproductive</td>
<td>LEP, MEST, UBB, IGFBP4, RETSAT, GSTM3, SERPINB2</td>
</tr>
<tr>
<td>Motor</td>
<td>FHL1, HBB, HBA1</td>
</tr>
<tr>
<td>Urinary System</td>
<td>UBC, GPX3, SOD1, NNMT, GEM, CRISPLD2</td>
</tr>
<tr>
<td>Nervous System</td>
<td>MMD, EIF4A2, BEX3, RTN4, NDUFS5, DYNLL1, UGP2, CALM2, APOD, ABCA8, RPS10, S100A10, RPL13, ARL6IP5</td>
</tr>
<tr>
<td>Skin/Skeletal Muscle System</td>
<td>FABP4, LPL, SCD, CYB5R3, CHRDL1, PCOLCE2, CIDE, ADIPOQ</td>
</tr>
<tr>
<td>Circulatory System</td>
<td>EFEMP1, SPARC, UQCR10, RPLP1, GPX4</td>
</tr>
</tbody>
</table>

Fig. 1. GSE5090 dataset after normalization.

Fig. 2. Differentially expressed genes (DEGs) between polycystic ovary syndrome (PCOS) and controls: (A) GSE5090 volcano plot; 17 significantly expressed genes were identified. Red, green and grey dots represent up-regulated, down-regulated and unchanged genes, respectively. (B) Heatmap of the top 50 DEGs of GSE5090.
3.5. Construction of PCOS Risk Prediction Columnar Chart Model

We then constructed a column-line graph model to predict the risk of PCOS prevalence (Figure 7A). Thus, our column-line graph model performed well in PCOS prevalence prediction. Subsequently, we calculated the ROC curves for the three pivotal genes (ANAX1, LEP, and GPX1) to evaluate their predictive performance.
Therapeutic targets in polycystic ovary syndrome genes

and RPL13) to assess the diagnostic effect (Figure 7B, C, and D).

4. Discussion

Little is known about the pathogenesis of Polycystic Ovary Syndrome (PCOS), making the exploration of its molecular mechanisms crucial for diagnosis and treatment. Research has identified certain genes that may be directly involved in regulating follicular growth, development, hormone synthesis, and metabolism in PCOS patients [7,8]. The identification of these differential genes could provide new insights into the diagnosis, treatment, and prognosis of the disease. Increasing evidence suggests that microRNAs (miRNAs) play significant roles in various human diseases, influencing processes such as development, proliferation, cell differentiation, cell cycle control, and cell death [9]. In this study, we screened 17 differentially expressed genes from omental adipose tissue of PCOS and control samples by analysing the differentially expressed genes from the GSE5090 data retrieved from the GEO database, and all of them were up-regulated in the PCOS samples. Three key genes were identified between hub genes and tissue-specific genes, including RPL13, LEP and ANXA1.

Ribosomes, the organelles that catalyze protein synthesis, consist of a small 40S subunit and a large 60S subunit. These subunits collectively comprise four RNAs and approximately 80 structurally distinct proteins. The RPL13 gene encodes a ribosomal protein that is a component of the 60S subunit. This protein is part of the L13E family of ribosomal proteins and is located in the cytoplasm. Notably, the expression level of this gene is significantly higher in benign breast lesions compared to breast cancer. Furthermore, RPL13 is highly expressed in the uterus and ovaries of females, leading to the hypothesis that RPL13 plays a major role in ovarian function [10].

The LEP gene encodes a protein that is secreted into the circulation by white adipocytes, playing a pivotal role in regulating energy homeostasis. Additionally, this protein serves multiple endocrine functions, being involved in the regulation of immune and inflammatory responses, haematopoiesis, angiogenesis, reproduction, bone formation, and wound healing. Notably, mutations in this gene and its regulatory region have been linked to severe and morbid obesity, accompanied by hypogonadism in human patients [11].

As a growth factor in certain tissues, this protein upregulates the expression of genes involved in cell cycle regulation by activating various signaling pathways. Examples include the upregulation of CCND3 through the JAK1-STAT2 pathway, and VEGFA via the MAPK1/3 and PI3K-AKT1 pathways. It also promotes the expression of genes that regulate the cell cycle. This is achieved by promoting the synergy between NOS1 and IL2B, primarily through the activation of their respective pathways. Additionally, it involves the interaction of JAK6, PI8K, and MAP2K2/VEGFA, which are integral to this process. VEGFA also acts as a pro-inflammatory agent, facilitating the cooperation between NOS1 and IL2B. This interaction further stimulates the production of IL1, IL14, and prostaglandin E38 via signaling pathways involving JAK6, PI8K, MAP2K2/MEK3, and MAP2K2/p1 [12,13].

This gene encodes a membrane-localized protein that binds phospholipids and exhibits anti-inflammatory activity by inhibiting phospholipase A2. A loss of function or
expression of this gene has been observed in various tumors. The protein plays a crucial role in the innate immune response, acting both as an effector of glucocorticoid-mediated responses and as a regulator of inflammatory processes. Its role extends to the down-regulation of the early stages of glucocorticoid-mediated inflammatory response (by similarity). Additionally, it promotes adaptive immune responses by enhancing the signaling cascade triggered by T cell activation and regulates the differentiation and proliferation of activated T cells. The protein specifically promotes T cell differentiation into Th1 cells while negatively regulating differentiation into Th2 cells, without affecting unstimulated T cells. It also negatively regulates hormone cytotoxicity through activation of the formyl peptide receptor and the reorganization of the actin cytoskeleton [14-17]. ANXA1 plays a critical role in the innate immune response, functioning both as an effector of glucocorticoid-mediated responses and a regulator of inflammatory processes, with notable anti-inflammatory activity [14]. It contributes to the down-regulation of the early stages of glucocorticoid-mediated inflammatory responses (by similarity). Additionally, ANXA1 regulates the differentiation and proliferation of activated T cells, thereby promoting adaptive immune responses by enhancing the signaling cascade triggered by T cell activation [16]. As a member of the class of calcium-phospholipid-binding proteins found widely in cells, ANXA1 regulates essential biological processes such as cellular signaling, differentiation, and apoptosis, primarily through its interaction with FPR1/FPR2. It can induce AKT phosphorylation by activating PI3K. Recent studies have highlighted ANXA1's association with the development of ovarian cancer, polycystic ovary syndrome, and chemotherapy resistance. These findings suggest its potential as a biomarker for drug resistance and prognosis in ovarian cancer [18,19].

However, this study has certain limitations. Firstly, it relies on gene chip data from GSE5090, which comprises a relatively small sample size. Secondly, since the study identified key genes based on different sample sources for Polycystic Ovary Syndrome (PCOS), validation across populations presents challenges. Consequently, future research should aim to increase the sample size, incorporate multicenter data, and conduct more population-based experiments. Such efforts are essential to validate findings in both PCOS patients and control groups. This expanded approach will enable a deeper exploration of the molecular mechanisms and pathways involved in PCOS. Ultimately, it may lead to the identification of new, valuable predictive biomarkers and the development of scientific and effective therapeutic options for PCOS diagnosis.

Conflict of Interests
The author has no conflicts with any step of the article preparation.

Consent for publications
The author read and approved the final manuscript for publication.

Ethics approval and consent to participate
No human or animals were used in the present research.

Informed Consent

The authors declare not used any patients in this research.

Availability of data and material
The data that support the findings of this study are available from the corresponding author upon reasonable request.

Authors’ contributions
YL and HM designed the study and performed the experiments, JY and JL collected the data, JY, JL and WL analyzed the data, YL and HM prepared the manuscript. All authors read and approved the final manuscript.

Funding
This work was supported by the National Natural Science Foundation of China (82104915), Project of Improving Basic Scientific Research Ability of Young and Middle-aged Teachers in Universities of Guangxi (2022KY0287).

References

13. Vuolteenaho K, Koskinen A, Kuukkonen M, Nieminen R, Paiava-