Research article

Bioinformatics analysis of the role of RNA modification regulators in polycystic ovary syndrome

Kewei Quan [a,b,*], Shuting Ning [c], Zilin You [a], Gaopi Deng [d]

[a] Department of Obstetrics and Gynecology, Dongguan Hospital of Guangzhou University of Chinese Medicine, Dongguan, 523000, China
[b] The First Clinical Medical College of Guangzhou University of Chinese Medicine, Guangzhou, 510405, China
[c] Department of Gynaecology and Obstetrics, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou, 510120, China
[d] Department of Gynecology, The First Affiliated Hospital of Guangzhou University of Chinese Medicine, Guangzhou, 510080, Guangdong Province, China

A B S T R A C T

Purpose: Polycystic ovary syndrome (PCOS) is the most common metabolic and endocrine disorder affecting women of reproductive age. The pathogenesis of PCOS is influenced by factors such as race, genetics, environment, hyperandrogenemia, hyperinsulinemia, and obesity. However, the molecular mechanisms linking RNA modification and PCOS remain underexplored. This study aims to investigate the potential genetic and molecular pathways connecting RNA modification with PCOS through bioinformatics analyses.
Methods: The GSE34526, GSE5850, and GSE98421 datasets were obtained from the National Center for Biotechnology Information Gene Expression Omnibus database. We identified intersecting differentially expressed genes (DEGs) and RNA modification-related genes within the GSE34526 dataset and visualized the overlaps using a Venn diagram. Subsequent analyses included Gene Ontology (GO), pathway enrichment (Kyoto Encyclopedia of Genes and Genomes), gene set enrichment analysis (GSEA), gene set variation analysis (GSVA), and immune infiltration analysis. Additionally, we constructed a protein-protein interaction network as well as mRNA-miRNA, mRNA-RNA binding protein, and mRNA-transcription factor (TF) regulatory networks. The expression and receiver operating characteristic curves of hub genes were also identified.
Results: The expression of several RNA modification-related DEGs (RMRDEGs) (*ALYREF, NUDT1, AGO2, TET2, YTHDF2,* and *TRMT61B*) showed significant differences in PCOS patients. GSEA and GSVA indicated that RMRDEGs were enriched in the hedgehog, MAPK, JAK STAT, and Notch pathways. Key transcription factors, including SP7, KLF8, HCFC1, IRF1, and MLLT1, were identified in the TF regulatory networks.
Conclusions: These findings suggest that there are gene and miRNA profile alterations exist in PCOS patients and highlight immune-related differences. This knowledge could pave the way for new research directions in the diagnosis and treatment of PCOS.

* Corresponding author. Department of Obstetrics and Gynecology, Dongguan Hospital of Guangzhou University of Chinese Medicine, Dongguan, 523000, China.
  E-mail address: 279271961@qq.com (K. Quan).

1. Introduction

Polycystic ovary syndrome (PCOS) is a prevalent endocrine disorder affecting women of reproductive age, characterized by symptoms such as irregular menstruation, acne, hirsutism, polycystic ovaries, amenorrhea, and infertility [1]. It is a leading cause of anovulatory infertility, impacting 8–13 % of women in this age group globally [2]. Although the pathogenesis of PCOS is not fully understood, various genetic, environmental, and hormonal factors have been implicated [3]. PCOS is also associated with several other conditions, including metabolic syndrome, type 2 diabetes, hypertension, cardiovascular diseases, and ovulatory infertility [4]. Despite advancements in the diagnosis and management of PCOS, the underlying molecular mechanisms and signaling pathways remain poorly understood.

Recent research has underscored the significance of post-transcriptional RNA modifications in the development of PCOS. These RNA modifications can influence epigenetic programming and gene expression [5]. Studies have demonstrated differences in DNA methylation and miRNA expression in women with PCOS [6]. MicroRNAs, which are small RNA molecules that regulate gene expression, have been found to play a role in ovarian tissues, promoting oocyte maturation and delaying reproductive aging [7,8]. Dysregulation of RNA modifications may contribute to PCOS development [9].

Although several studies have investigated PCOS biomarkers and epigenetic regulation, research on RNA modifications in PCOS is still lacking. This study aims to explore RNA modification biomarkers and their associated cellular pathways in PCOS. Additionally, the relationship between RNA modifications and immune cell infiltration is examined to identify potential targets for immunotherapy. By analyzing gene expression and conducting enrichment analysis, we aim to provide a deeper understanding of the pathological processes and signaling pathways involved in PCOS.

2. Methods

2.1. Data sources

Transcriptomic and clinical data of PCOS patients was downloaded from the Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/) database [10] with R package GEOquery [11].The datasets included were GSE34526 [12], GSE5850 [13], and GSE98421 [14].GSE34526 consisted of 10 samples, with seven PCOS patients and three controls. GSE5850 included six PCOS samples and six controls, while GSE98421 comprised four PCOS samples and four controls. GSE34526 served as the test set, while GSE5850 and GSE98421 acted as verification sets.

During dataset retrieval and preliminary screening, we used the keywords "PCOS" and "gene expression" to search relevant datasets in NCBI GEO database, collecting more than 700 samples. To ensure the consistency and comparability of the data, we selected datasets that utilized the same gene expression detection platform. In this study, we chose datasets from the GPL570 [HG-U133_Plus_2] platform, as it is widely used in PCOS research and provides reliable data quality. Furthermore, we selected human datasets to ensure direct clinical relevance of our findings.

We evaluated the quality of all datasets meeting the platform and species criteria, excluding those with significant missing data or incomplete labeling. Among the selected datasets, we prioritized those with moderate sample sizes and clearly labeled health and PCOS groups. To ensure balance between groups, we selected datasets where the number of samples in the healthy group and the PCOS group were relatively equal. Ultimately, we selected three datasets—GSE34526, GSE5850 and GSE98421. Although the total sample size was 30, these datasets were chosen based on strict criteria, including platform consistency, species consistency, sample quality, and balance between groups, which greatly improved the accuracy and reliability of our analysis results.

We identified 95 RNA modification-related genes (RMRGs) from relevant literature [15–17]. Details of the datasets are listed in Table 1. Further details on the identified RMRGs are provided in Table S1.

2.2. Identification of PCOS-associated DEGs

To elucidate the mechanisms and pathways involved in PCOS, we utilized the limma package [18] to identify differentially expressed genes (DEGs) from microarray data in GSE34526, GSE5850, and GSE98421. Thresholds of $p < 0.05$ and $|logFC|>0$) were applied, allowing for a more comprehensive capture of potentially significant genes, as seen in prior studies (e.g., PMID: 38253683 and PMID: 33144895).DEGs were filtered using these criteria: genes with $|logFC| > 0$ and $p < 0.05$ were considered upregulated, while those with $|logFC| < 0$ and a $p < 0.05$ were considered downregulated. To identify PCOS-associated RNA modification-related DEGs

Table 1

Polycystic ovary syndrome (PCOS) dataset information list.

	GSE34526	GSE5850	GSE98421
Platform	GPL570	GPL570	GPL570
Species	*Homo sapiens*	*Homo sapiens*	*Homo sapiens*
Tissue	Granulosa cells	/	Fat biopsy
Samples in the Normal group	3	6	4
Samples in the PCOS group	7	6	4
Reference	PMID: 22904171	PMID: 17148555	/

(RMRDEGs), we compared DEGs with RNA modification-related genes (RMRGs) in GSE34526 and visualized the overlaps using a Venn diagram, volcano map (ggplot2), and heatmap (pheatmap) in R.The reliability of limma in differential expression analysis, combined with the powerful visualization capabilities of ggplot2, ensures robust data analysis.

2.3. Functional (GO) and pathway enrichment (KEGG) analyses

Gene Ontology (GO) [19] and Kyoto Encyclopedia of Genes and Genomes (KEGG) [20] pathway analysis were performed using the R package clusterProfiler [21], with GO annotation analysis specifically conducted on RMRDEGs. The advantage of clusterProfiler is its comprehensive range of functional enrichment analysis options. Statistical significance was evaluated using the Benjamini–Hochberg method. Enrichment analysis was conducted with a significance threshold of $p < 0.05$ to identify significant genes in the pathological group compared to controls.

2.4. Gene set enrichment analysis

We performed Gene Set Enrichment Analysis (GSEA) [22] to evaluate the tendency of genes within predefined gene sets to be distributed in a ranked gene list by phenotypic correlation, and thus determining contribution to the phenotype. The genes in the dataset GSE34526 were first divided into high and low groups. Using the "clusterProfiler" package in R, we conducted GSEA with parameters including a seed number of 2020, 1000 permutations, gene set containing 10 and 500 genes, and the Benjamini–Hochberg P-value correction method. We obtained the gene set "H.A.v7.4. Symbols.gmt" from the Molecular Signatures Database (MSigDB) for GSEA analysis in GSE34526, calculating the functional enrichment differences of DEGs between normal and the PCOS samples. Significant enrichment was determined with $p < 0.05$ and a false discovery rate (FDR) < 0.05.

2.5. Gene set variation analysis

Gene Set Variation Analysis (GSVA) [23] was employed to assess gene set enrichment in the microarray nuclear transcriptome of the GSE34526 dataset.Using the "H.A.v7.4. symbols.gmt" gene set from the MSigDB database, we investigated the functional enrichment differences between normal and PCOS patient samples.

2.6. Immune infiltration analysis

We uploaded the gene expression matrix to the CIBERSORTx [24] online tool (https://cibersortx.stanford.edu/), integrated with the *Homo sapiens* gene matrix, and identified immune cell enrichment scores greater than zero in the abundance matrix. Differences in immune cell correlations among groups were calculated using the Wilcoxon test, with significance set at $p < 0.05$. Correlations between different immune cells were assessed using Spearman's rank correlation coefficient and visualized with ggplot2. Additionally, we merged the gene expression matrix of the PCOS dataset to compute correlations between immune cells and RNA-modified DEGs, creating a correlation heatmap displayed using the pheatmap package.

2.7. Protein-protein interaction network construction

A protein-protein interaction (PPI) network was constructed to depicts interacting proteins involved in various biological processes, providing insights into disease pathogenesis and progression. We used the STRING database [25] to construct a PPI network associated with DEGs from screened RMRDEGs, setting criteria for interaction score and the maximum number of interactors. The network was used to analyze and visualize using Cytoscape [26] (version 3.9.1), identifying Key RMRDEGs serving as hub genes for PCOS.

2.8. Construction of the mRNA-miRNA, mRNA-RBP, and mRNA-TF regulatory network

Using the ENCORI database (http://starbase.sysu.edu.cn/) (version 3.0) [27],we explored miRNA and RNA interactions based on CLIP-seq and biodegradation sequencing data. The miRDB database [28] was utilized to predicts miRNA target genes, which helped construct the mRNA-miRNA interaction network. Additionally, the ENCORI database predicted interactions between RNA-binding proteins (RBPs) and hub genes (mRNA), forming the mRNA-RBP interaction network. The CHIPBase database (version 3.0) [29] and the TFtarget database were employed to identify transcription factors (TFs) and their regulatory regulations with hub genes (mRNA).

2.9. Receiver operating characteristic curves and expression of hub genes

In the GSE34526 dataset, correlation heatmaps were generated to examine the relationship between the gene expression levels of RNA modification-related DEGs in PCOS and identify correlations among hub genes. Receiver operating characteristic (ROC) curves were plotted using the pROC package to evaluate the diagnostic effectiveness of hub gene expression in different subgroups (normal/PCOS) across the three datasets. The area under the curve (AUC) was calculated to assess the diagnostic performance of hub gene expression, with AUC values were categorized as low accuracy (0.5–0.7), certain accuracy (0.7–0.9), and high accuracy (>0.9).

2.10. Statistical analysis

Statistical analyses were performed using RStudio (version 4.1.2). The Student's *t*-test and false discovery rate (FDR) were applied to determine statistical significance in differential expression analysis. The independent Student's *t*-test, Wilcoxon test, and Chi-square test or Fisher's exact test were used to assess normally distributed variables, continuous variables, and categorical variables, respectively. Correlation coefficients were calculated using Spearman's correlation analysis. All P-values were two-tailed, with significance was set at $P < 0.05$.

3. Results

We employed bioinformatics methods to explore the biological characteristics of PCOS. A schematic of the study is shown in Fig. 1. The research process began with the gene expression dataset GSE34526. Initially, GSEA and DEG analysis were performed. During the analysis of DEGs, genes related to RNA modification were identified and integrated with the DEGs to obtain RMRDEGs. Subsequently, a PPI network was constructed to analyze the RMRDEGs and identify key genes or proteins (hub genes).

Further analysis of the expression differences in these key genes was conducted, followed by ROC curve analysis of other gene expression datasets (GSE5850 and GSE98421) to evaluate the predictive capability of the model.Additionally, results from the PPI network analysis were utilized for further bioinformatics research, including the analysis of microRNA regulatory effects on gene expression (mRNA-miRNA), the impact of RNA-binding proteins on gene expression (mRNA-RBP), and the regulation of gene expression by TFs.

GO and KEGG pathways were performed to understand the functions and interaction pathways of genes. Finally, analyses related to the immune system were performed. This comprehensive research process was designed to provide a deeper understanding of the mechanisms, functions, and interactions of gene expression regulation, as well as biological processes (BPs) related to the immune system.

3.1. Data processing results

After normalizing the datasets GSE34526, GSE5850, and GSE98421, the expression levels across different samples tended to be consistent, as illustrated in Fig. S1.

Fig. S1 displays box plots for three different gene expression datasets: GSE34526, GSE5850, and GSE98421. Each dataset is represented by two figures: the box plot before normalization (on the left, labeled A, C, E) and the box plot after normalization (on the right, labeled B, D, F). Each figure includes distribution box plots for multiple samples.

A) "GSE34526 Before Normalization" shows the distribution of gene expression values prior to normalization. Each box plot represents the distribution of gene expression in a sample, where the "box" indicates the median (the middle horizontal line within the
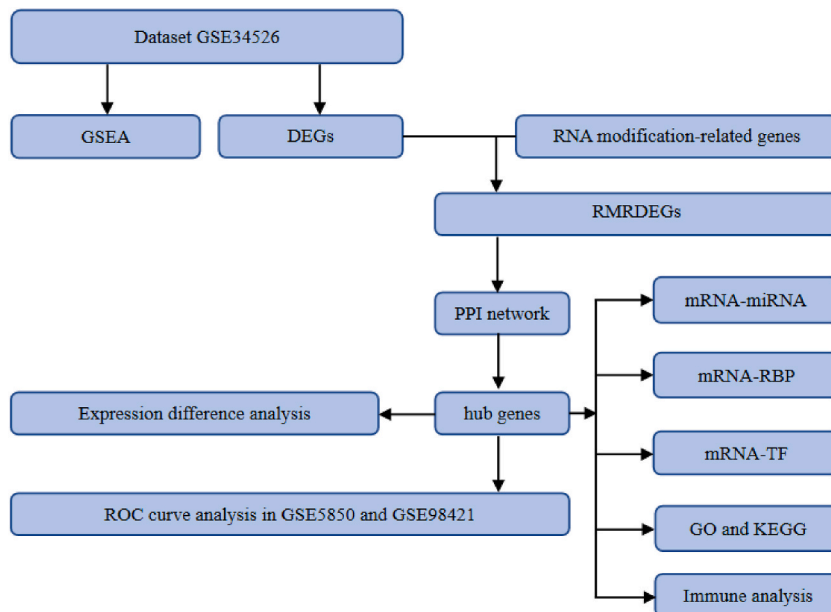

Fig. 1. Flowchart showing the progression of the study
PCOS: Polycystic ovary syndrome; RMRDEGs: RNA modification-related genes that were differentially expressed; GSEA: Gene set enrichment analysis; GSVA: Gene set variation analysis; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; PPI: Protein–protein interaction; TF: Transcription factor; RBP: RNA-binding protein; DEGs: Differentially expressed genes; ROC: Receiver operating characteristic.

box) and the upper and lower quartiles (the top and bottom of the box), and the "whiskers" represent the maximum and minimum values, as well as potential outliers (not explicitly indicated in the figure). B) "GSE34526 After Normalization" presents the distribution of gene expression values for the same dataset GSE34526 after normalization. Compared to the pre-normalization figure, the values here are more standardized and consistent, reducing individual variation and technical bias, making the gene expression data more comparable. C) "GSE5850 Before Normalization" illustrates the distribution for the GSE5850 dataset before normalization, while D) "GSE5850 After Normalization" displays the sample distribution after normalization. E) "GSE98421 Before Normalization" and F) "GSE98421 After Normalization" represent the box plots for the GSE98421 dataset before and after normalization, respectively.

In each set of figures, the range of expression values in the box plots after normalization appears more similar, reflecting better consistency in the data after processing. Normalization is a crucial step in bioinformatics analysis, providing standardized data for further analysis. Each box plot represents the distribution of expression values across all genes for a sample.

3.2. Analysis of DEGs associated with PCOS

Using the limma package, we analyzed the GSE34526, GSE5850, and GSE98421 datasets to identify DEGs associated with PCOS. In the GSE34526 dataset, 2077 genes met the criteria, with 1320 upregulated and 757 downregulated in the PCOS group. The volcano map (Fig. 2A) represents these results. Similarly, the GSE5850 dataset displayed 1113 threshold genes, with 485 upregulated and 628 downregulated (Fig. 2B). For the GSE98421 dataset, 567 threshold genes were identified, with 245 upregulated and 322 down-regulated (Fig. 2C).

To identify RMRDEGs, we analyzed the GSE34526 dataset, revealing six RMRDEGs (ALYREF, NUDT1, AGO2, TET2, YTHDF2, TRMT61B). A Venn diagram (Fig. 2D) displays the different groups (normal/PCOS) in the GSE34526 dataset, while a heatmap (Fig. 2E) represents the clustering of these RMRDEGs. RNA modification information for these six DEGs is shown in Table S2 of the supplementary document.

In Fig. 2: A–C) are the Volcano plots used to display the results of differential gene expression analysis. The -log10(p-value) is plotted on the y-axis to indicate significance and the log2 fold change is plotted on the x-axis to illustrate changes in expression levels.
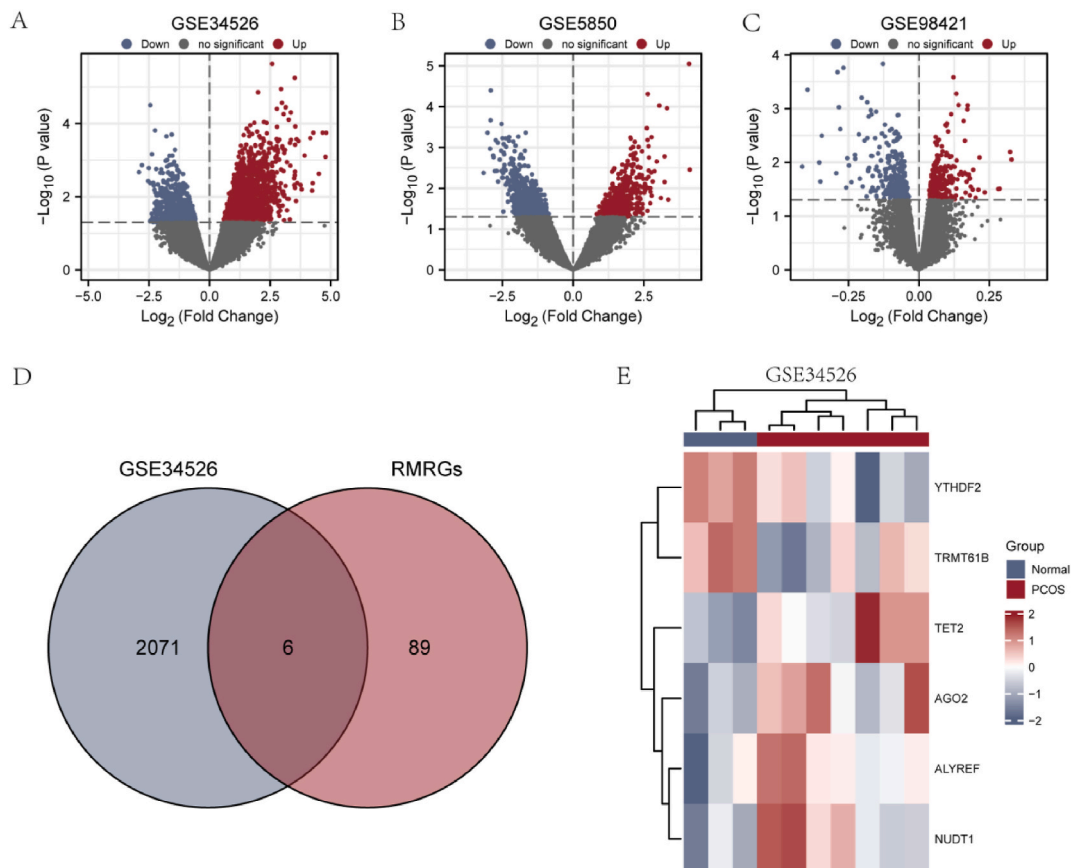

Fig. 2. Differential gene analysis in the PCOS dataset
A. Differential gene Volcano map of the GSE34526 dataset. B. Differential gene volcano map of the GSE5850 dataset. C. Differential gene volcano map of the GSE98421 dataset. D. Venn diagrams of DEGs and RMRGs in the data set. DEGs: Differentially expressed genes; RMRGs: RNA modification related genes; PCOS: Polycystic ovary syndrome.

In these plots, the red dots represent significantly upregulated genes, blue dots represent significantly downregulated genes, and gray dots indicate genes without significant changes. Each plot corresponds to a different gene expression dataset (GSE34526, GSE5850, GSE98421). D) is a Venn diagram for comparing overlapping and unique elements between two data sets. Here, genes in the GSE34526 dataset are compared with those in RMGs (which may refer to a specific type of genes or genomes). It shows that there are six genes common between the two, 2071 unique genes in GSE34526, and 89 unique genes in RMGs. E) is a hierarchical heatmap showing variations in gene expression levels across samples. The dendrogram above the heatmap indicates clustering among samples or genes. Each row in the heatmap represents a gene, and each column represents a sample, with color intensity reflecting the level of gene expression (blue for lower expression, red for higher expression). Labels on the right side indicate different sample groups helping identify which genes exhibit significant differences in expression under different states or disease conditions.

3.3. Functional enrichment analysis (GO) and pathway enrichment (KEGG) of DEGs associated with RNA modification

Functional enrichment analyses were conducted to understand the biological pathways and relationships of the six identified RMRDEGs: ALYREF, NUDT1, AGO2, TET2, YTHDF2, and TRMT61B in PCOS. GO (Table 2) and KEGG (Table 3) enrichment analyses revealed significant findings. Enrichment items with a P-value <0.05 and FDR value < 0.05 were considered statistically significant. The results are presented using bubble diagrams (Fig. 3A and B) and ring network diagrams (Fig. 3C and D).

The six RMRDEGs are mainly enriched in the following biological processes (GO-BP): non-canonical Wnt signaling pathway, regulation of Notch signaling pathway, And RNA methylation. In terms of cellular components (GO-CC), they are mainly concentrated in mitochondrial matrix, ribonucleoprotein granule, acrosomal vesicle and polysome. In terms of molecular function (GO-MF), they involve endonuclease activity, catalytic activity (acting on a tRNA), ribonuclease activity, and single-stranded RNA binding, etc.KEGG analysis revealed that these genes are mainly involved in the following pathways: Amyotrophic lateral sclerosis,RNA transport, Spliceosome, and mRNA surveillance pathway.

In Fig. 3: A) is a scatter plot for Gene Ontology (GO) enrichment analysis, depicting enrichment in the BPs, cellular components (CCs), and molecular functions (MFs) associated with a set of genes. The vertical axis shows different GO categories, while the horizontal axis represents the gene ratio, which is the ratio of the number of significantly enriched genes in a specific GO category to the total number of genes in that category. The size of the circles indicates the number of genes in the category (Counts), and the color intensity reflects the corrected P-value (p.adjust), with darker colors indicating higher statistical significance. B) is a scatter plot for KEGG pathway enrichment analysis, showcasing the enrichment of different pathways in a manner similar to the GO analysis. The vertical axis displays various KEGG pathways, such as the "mRNA surveillance pathway" and "Spliceosome," with the horizontal axis also representing the gene ratio. Circle size denotes the number of genes in the pathway category, while the color signifies the level of significance. C) is a gene-function network diagram that connects genes (on the left, such as *NUDT1* and *AGD2*) with GO functional categories (on the right), with straight lines illustrating the associations between genes and their related functional categories. The size of the circles indicates the number of associated genes, though no explanation for the colors is provided. D) is a gene-KEGG pathway network diagram. It is similar to the GO network diagram; however, it shows the relationships between genes (on the left) and KEGG pathways (on the right). For example, the gene *ALYREF* is associated with the KEGG pathway hsa03015. Lines indicate connections, and circle size reflects the number of associated genes.

3.4. GSEA and GSVA enrichment analysis of the PCOS dataset

GSEA and GSVA enrichment analyses were conducted on the PCOS dataset (GSE34526) to assess the impact of DEG expression levels (Fig. 4). The relationship between DEGs and their associated biological processes (BPs) in different groups (normal/PCOS) was

Table 2
GO enrichment analysis results of RNA modification related DEGs.

Ontology	ID	Description	GeneRatio	BgRatio	P-value	p.adjust	Q-value
BP	GO:0061014	Positive regulation of the mRNA catabolic process	2/6	48/18670	9.65e-05	0.021	0.008
BP	GO:1903313	Positive regulation of the mRNA metabolic process	2/6	78/18670	2.56e-04	0.021	0.008
BP	GO:0006446	Regulation of translational initiation	2/6	79/18670	2.62e-04	0.021	0.008
BP	GO:2000241	Regulation of the reproductive process	2/6	146/18670	8.92e-04	0.033	0.013
BP	GO:0006413	Translational initiation	2/6	193/18670	0.002	0.033	0.013
CC	GO:0000932	P-body	2/6	84/19717	2.66e-04	0.006	0.002
CC	GO:0036464	Cytoplasmic ribonucleoprotein granule	2/6	212/19717	0.002	0.014	0.004
CC	GO:0035770	Ribonucleoprotein granule	2/6	223/19717	0.002	0.014	0.004
CC	GO:0005845	mRNA cap binding complex	1/6	12/19717	0.004	0.016	0.004
CC	GO:0000346	Transcription export complex	1/6	14/19717	0.004	0.016	0.004
MF	GO:0008174	mRNA methyltransferase activity	1/6	11/17697	0.004	0.035	0.007
MF	GO:0035197	siRNA binding	1/6	11/17697	0.004	0.035	0.007
MF	GO:0047429	Nucleoside-triphosphate diphosphatase activity	1/6	11/17697	0.004	0.035	0.007
MF	GO:0000340	RNA 7-methylguanosine cap binding	1/6	12/17697	0.004	0.035	0.007
MF	GO:0000339	RNA cap binding	1/6	19/17697	0.006	0.035	0.007

RMRDEGs: RNA modification-related differentially expressed genes; GO: Gene Ontology; BP: Biological process; CC: Cellular component; MF: Molecular function.

Table 3
KEGG enrichment analysis results of RNA modification related DEGs.

Ontology	ID	Description	GeneRatio	BgRatio	P-value	p.adjust	Q-value
KEGG	hsa03015	mRNA surveillance pathway	1/1	97/8076	0.012	0.038	0.008
KEGG	hsa03040	Spliceosome	1/1	151/8076	0.019	0.038	0.008
KEGG	hsa03013	RNA transport	1/1	186/8076	0.023	0.038	0.008
KEGG	hsa05014	Amyotrophic lateral sclerosis	1/1	364/8076	0.045	0.056	0.012
KEGG	hsa05168	Herpes simplex virus 1 infection	1/1	498/8076	0.062	0.062	0.013

KEGG: Kyoto Encyclopedia of Genes and Genomes; RMRDEGs: RNA modification-related differentially expressed genes.
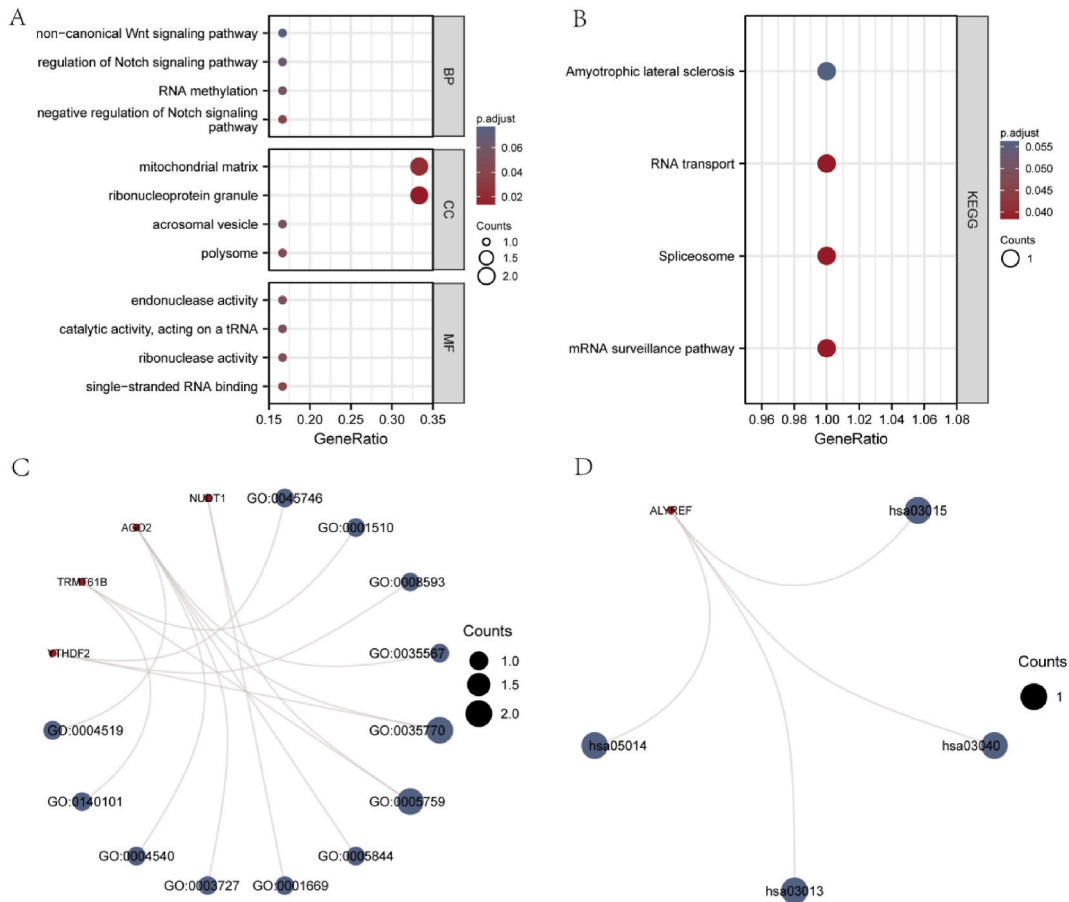

Fig. 3. RMRDEG functional enrichment analysis (GO) and pathway enrichment analysis (KEGG)
Functional enrichment analysis (GO) (A, C) and KEGG pathway enrichment analysis results (B, D) of RMRDEGs. RMRDEGs: RNA modification related genes differentially expressed genes; GO: Gene Ontology; BP: Biological process; CC: Cellular component; MF: Molecular function; KEGG: Kyoto Encyclopedia of Genes and Genomes. The screening criteria for GO and KEGG enrichment entries were a P-value <0.05 and false discovery rate (Q-value) <0.05.

examined. The DEGs exhibited significant enrichment in pathways such as Hedgehog (Fig. 4B), MAPK (Fig. 4C), JAK-STAT (Fig. 4D), and Notch (Fig. 4E–Table 4). Furthermore, GSVA enrichment analysis was performed on the GSE34526 dataset to compare PCOS and normal samples (Fig. 4F). This analysis revealed significant differences in functional enrichment between the two groups, particularly in pathways such as Notch, Wnt, IL2, and IL6 (Table 5).

In Fig. 4: A) is a density distribution plot from the gene set enrichment analysis (GSEA) highlighting the distribution of gene expression data for four cellular signaling pathways—Hedgehog, MAPK, JAK-STAT, and Notch—in the dataset GSE34526. From left to right, the plots show the variation in expression levels of these genes within the dataset from low to high, aiding in identifying differences in pathway activity.

B–E): These are enrichment plots from the GSEA, corresponding to the analysis results for the Hedgehog, MAPK, JAK-STAT, and Notch signaling pathways, respectively. The top curve of each plot demonstrates how the enrichment score (ES) varies with the ranked gene sets. The ES indicates the tendency of a gene set to cluster toward the top or bottom of the dataset. The bottom bars (black vertical
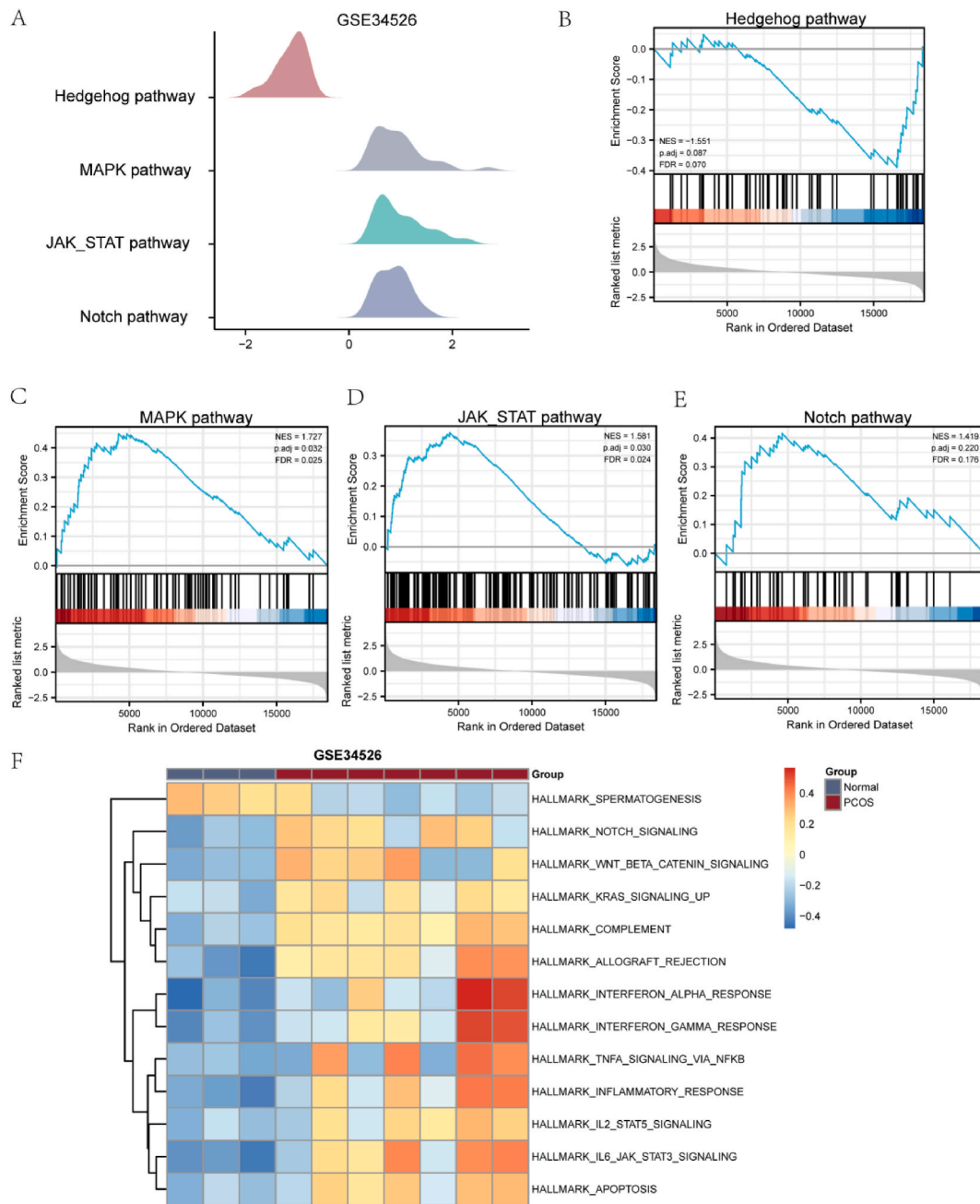
Fig. 4. GSEA and GSVA enrichment analysis of the GSE34526 PCOS dataset
A. GSEA of the GSE34526 dataset. B–E. The differentially genes in the GSE34526 dataset were significantly enriched in the Hedgehog pathway (B), MAPK pathway (C), JAK STAT pathway (D), and Notch pathway (E). F. GSVA of the GSE34526 dataset. Blue represents the normal sample group and red represents the PCOS patient sample group. PCOS: Polycystic ovary syndrome; GSEA: Gene set enrichment analysis; GSVA: Gene set variation analysis. The screening criteria for GSEA and GSVA enrichment analyses were $P < 0.05$ and a false discovery rate value (P value) < 0.05.

lines) show the ranking of the specific gene set across the entire list of genes, with red indicating the highest trend of enrichment at that point. The plots also display the normalized enrichment score (NES) and the false discovery rate (FDR).

F): This heatmap represents the activity index of different signaling pathways across samples, indicating their gene set scores. Clustering analysis of the samples reveals significant differences between normal (Normal) and polycystic ovary syndrome (PCOS) samples in the expression activity of multiple signaling pathways. The color strip represents changes in the activity index, with blue indicating lower activity and red indicating higher activity.

Table 4
GSEA analysis of GSE34526 dataset.

Description	Set Size	Enrichment Score	NES	P-value	Q-value
KEGG_HEDGEHOG_SIGNALING_PATHWAY	53	−0.38933726	−1.551408026	0.013736264	0.069866662
BIOCARTA_MAPK_PATHWAY	80	0.448943803	1.727238945	0.002898551	0.025339617
KEGG_JAK_STAT_SIGNALING_PATHWAY	152	0.375749027	1.581280386	0.002649007	0.024018199
WP_NOTCH_SIGNALING	44	0.416721058	1.418660187	0.053398058	0.176443176
KEGG_CYTOKINE_CYTOKINE_RECEPTOR_INTERACTION	250	0.40029646	1.771594646	0.001221001	0.017365304
WP_MAPK_SIGNALING_PATHWAY	236	0.382955129	1.687913094	0.001226994	0.017365304
REACTOME_INTERFERON_SIGNALING	186	0.497277529	2.142536431	0.001272265	0.017365304
KEGG_CHEMOKINE_SIGNALING_PATHWAY	179	0.475981098	2.043156396	0.001282051	0.017365304
WP_VITAMIN_D_RECEPTOR_PATHWAY	164	0.438926203	1.865810445	0.00131406	0.017365304
WP_CHEMOKINE_SIGNALING_PATHWAY	161	0.461873975	1.950643205	0.001335113	0.017365304
WP_HEPATITIS_B_INFECTION	148	0.416609384	1.745983235	0.001338688	0.017365304
REACTOME_TOLL_LIKE_RECEPTOR_CASCADES	145	0.539202181	2.250308586	0.001355014	0.017365304
NABA_ECM_AFFILIATED	141	0.428886833	1.787868154	0.001358696	0.017365304
KEGG_NATURAL_KILLER_CELL_MEDIATED_CYTOTOXICITY	128	0.475032981	1.966855354	0.001362398	0.017365304
KEGG_CELL_ADHESION_MOLECULES_CAMS	125	0.442460334	1.827867032	0.00136612	0.017365304

GSEA: Gene set enrichment analysis.

Table 5
GSVA analysis results of GSE34526 dataset.

Description	logFC	AveExpr	t	P-Value	B
HALLMARK_IL6_JAK_STAT3_SIGNALING	0.573437762	−0.001917768	3.431651992	0.001563163	−1.155628174
HALLMARK_ALLOGRAFT_REJECTION	0.54725735	0.018321638	3.348258541	0.00196351	−1.358354491
HALLMARK_INFLAMMATORY_RESPONSE	0.502184956	−0.02553951	3.052933165	0.004322486	−2.056087693
HALLMARK_INTERFERON_GAMMA_RESPONSE	0.477406957	−0.024534411	2.852626222	0.007248781	−2.509152526
HALLMARK_COMPLEMENT	0.464929639	0.058184326	2.726586773	0.009951851	−2.784842811
HALLMARK_SPERMATOGENESIS	−0.407470319	−0.034641677	−2.695622036	0.010746389	−2.851391302
HALLMARK_INTERFERON_ALPHA_RESPONSE	0.478556216	−0.066120837	2.666408386	0.011549555	−2.913735465
HALLMARK_NOTCH_SIGNALING	0.422849057	−0.003580058	2.644791124	0.012179265	−2.959588996
HALLMARK_WNT_BETA_CATENIN_SIGNALING	0.408537465	−0.011178712	2.451810136	0.019375666	−3.357979318
HALLMARK_APOPTOSIS	0.396896132	0.010790873	2.411608018	0.021295095	−3.438393105
HALLMARK_IL2_STAT5_SIGNALING	0.35818706	−0.002127791	2.232680643	0.032101453	−3.784821298
HALLMARK_TNFA_SIGNALING_VIA_NFKB	0.391963242	−0.022104131	2.16194063	0.03758126	−3.916402428
HALLMARK_KRAS_SIGNALING_UP	0.319926245	−0.004284628	2.058993963	0.047037772	−4.102178567

GSVA: Gene set variation analysis.

3.5. Immunoinfiltration analysis of the PCOS dataset (CIBERSORTx)

We utilized the CIBERSORTx online tool to calculate the correlation between 22 immune cell types and the expression profile data of the GSE34526 PCOS dataset. The resulting immunoinfiltration analysis was visualized using a bar graph to display immune cell infiltration in each sample (Fig. 5A). Correlation heatmaps were generated to assess the relationships between the six RMRDEGs (ALYREF, NUDT1, AGO2, TET2, YTHDF2, and TRMT61B) and the abundance of the 22 immune cell types in the GSE34526 dataset (Fig. 5B). The results revealed correlations between the RMRDEGs and the abundance of 14 immune cell infiltrates. Notably, most immune cell types exhibited significant correlations with RNA-modified DEGs ($P < 0.005$).

The specific meanings of Fig. 5 is explained as follows: A) is a stacked bar chart representing the relative percentages of different cell types in normal and polycystic ovary syndrome (PCOS) samples. The cell types include various T cell subtypes, B cells, monocytes, natural killer (NK) cells, myeloid cells such as macrophages and dendritic cells, and other types of cells. This chart illustrates the distribution changes of each cell type, revealing shifts in the proportions of certain cell types within PCOS samples.

B) is a heatmap that displays the analysis results of the association between six genes (*ALYREF, NUDT1, AGO2, TET2, YTHDF2, TRMT61B*) and various cell types. In the heatmap, each row represents a gene, and each column represents a cell type. The color of the cell indicates the correlation of gene expression with the specific cell type, ranging from blue (negative correlation) to red (positive correlation). The deeper the color, the stronger the correlation. Additionally, the heatmap includes markers "*" and "**," denoting levels of statistical significance, indicating that the associations between these gene expressions and specific cell types are statistically significant.

3.6. PPI network: hub Gene-miRNA interaction network; hub Gene-RBP interaction network; hub Gene-TF interaction Network; Expression and correlation analyses of hub genes in the dataset

PPI analysis of the six RMRDEGs (*ALYREF, NUDT1, AGO2, TET2, YTHDF2, TRMT61B*) was performed using the STRING database. The resulting PPI network was visualized using Cytoscape software (Fig. 6A).

We applied mRNA-miRNA data from the ENCORI and miRDB databases to predict miRNAs that interact with the six hub genes. The
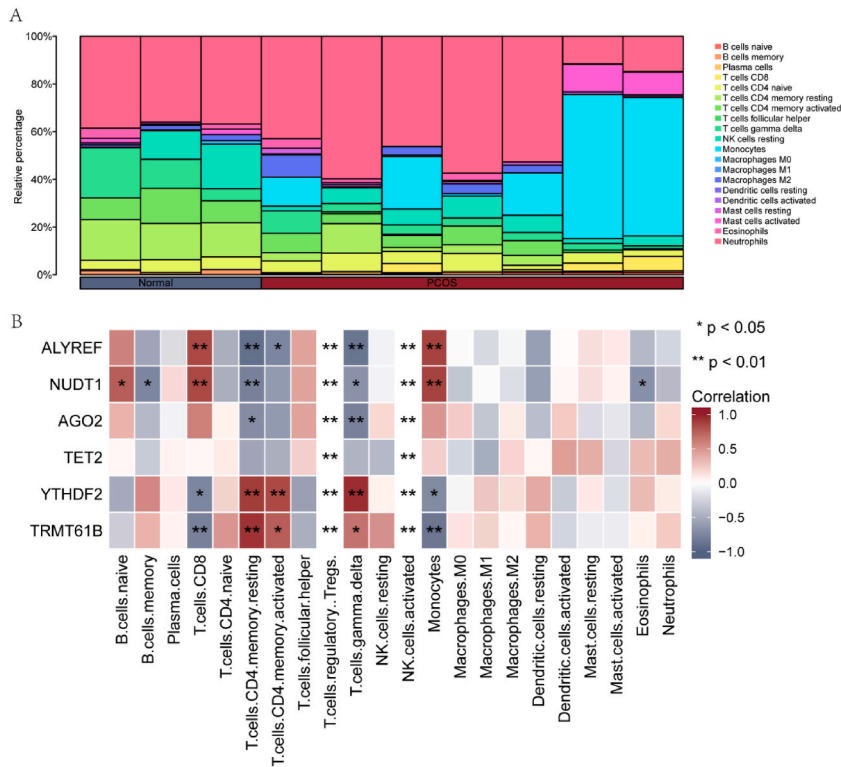
Fig. 5. Immunoinfiltration analysis of the GSE34526 dataset (CIBERSORTx) A. A histogram of the results of the immunoinfiltration analysis of 22 types of immune cells in the GSE34526 dataset. B. Heat map showing the correlation between RMRDEGs and immune cells.
"*" indicates a statistical significance of P < 0.05. "**" indicates a high statistical significance of P < 0.01. RMRDEGs: RNA modification-related genes that were differentially expressed.

intersection of results from both databases was extracted and visualized as an mRNA-miRNA interaction network using Cytoscape software (Fig. 6B). This network consisted of five hub genes (*NUDT1, AGO2, TET2, YTHDF2,* and *TRMT61B*), 179 miRNA molecules, and 195 mRNA-miRNA interaction pairs (Table S3).

RBP interactions were predicted using mRNA-RBP data from the ENCORI database. The mRNA-RBP interaction network, composed of five hub genes (*ALYREF, AGO2, TET2, YTHDF2, TRMT61B*) and 24 RBP molecules, was visualized using Cytoscape software (Fig. 6C). It consisted of 33 mRNA-RBP interaction pairs (Table S4). To identify TFs binding to hub genes, we searched the CHIPBase (version 3.0) and hTFtarget databases. The resulting interactive relationships and interactions with hub genes (mRNA) were downloaded and visualized using Cytoscape software. The network included six hub genes (ALYREF, NUDT1, AGO2, TET2, YTHDF2, and TRMT61B) and 77 TFs (Fig. 6D–Table S5).

The specific meanings of Fig. 6 are explained as follows:

A) : A protein-protein interaction (PPI) network diagram, where central nodes (such as *NUDT1, AGO2, TET2, YTHDF2*) represent core genes or proteins, forming complex interaction networks with surrounding genes. This indicates that these core nodes play significant roles in their biological processes.

B) : An mRNA-miRNA interaction network, centered around core nodes (TRMT61B, TET2, YTHDF2, NUDT1), illustrating interactions within the entire dataset.This network highlights the role of miRNAs in regulating the expression of these hub genes.

C) : An mRNA-RBP interaction network involving genes related to RNA splicing, modification, and metabolism (such as ALYREF, YTHDF2, and TET2).This network sheds light on the roles of these nodes in RNA biology.

D) : A transcription regulation network, incorporating transcription factors (such as STAT1, EP300, SPI1) and other proteins related to gene expression regulation. The interactions among these nodes provide insights into the complexity of gene expression regulation.

In the PCOS dataset (GSE34526), we examined the expression differences and correlations of the six RMRDEGs. The correlation analysis revealed significant differences in the expression levels of most hub genes (*ALYREF, NUDT1, AGO2, TET2, YTHDF2, TRMT61B*) between the normal and PCOS groups (Fig. 7A). Specifically, the expression levels of these hub genes showed statistically significant differences among the different groups (P < 0.05). To further investigate the correlation between the six hub genes (*ALYREF, NUDT1, AGO2, TET2, YTHDF2,* and *TRMT61B*), a heatmap displaying their correlations in the GSE34526 dataset was
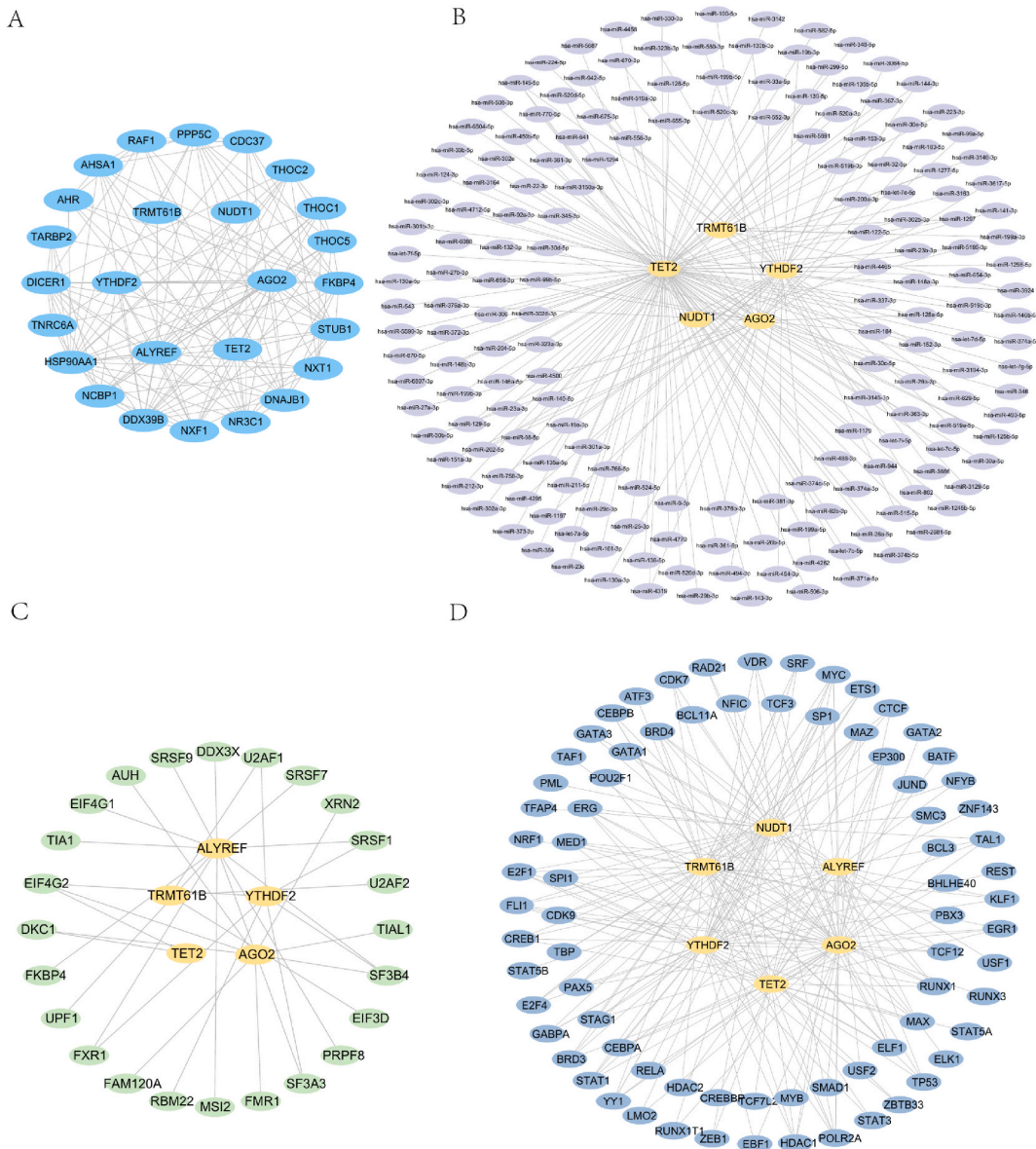
Fig. 6. Protein-protein interaction (PPI) network showing hub gene-miRNA interactions, hub gene-RBP interactions, and hub gene-TF interactions A. The PPI network of ALYREF, NUDT1, AGO2, TET2, YTHDF2, and TRMT61B. B. The hub gene-miRNA interaction network. C. The hub gene-RBP interaction network. D. The hub gene-TF interaction networks. Yellow ovals represent mRNA; purple ellipses represent miRNAs; green ellipses represent RBPs; dark blue ovals represent TFs. PPI network: Protein–protein interaction network; TF: Transcription factor.

generated (Fig. 7B). The results indicated statistically significant differences, particularly between ALYREF and NUDT1 (P < 0.01) and between ALYREF and YTHDF2 (P < 0.05).The following is a specific explanation of the meaning of Fig. 7:

A) Box plots displaying the distribution of expression levels for six specific genes (*ALYREF*, *NUDT1*, *AGO2*, *TET2*, *YTHDF2*, *TRMT61B*) in the dataset GSE34526. Blue box plots represent normal samples, and red box plots represent PCOS samples. The box in each box plot indicates the median (the horizontal line in the middle of the box) and the upper and lower quartiles (the edges of the box), while the "whiskers" denote the range of the data, excluding outliers. Small squares represent the mean value for each box plot. Star markers indicate significant differences between the two groups, with "*" denoting a p-value less than 0.05, and "ns" indicates a non-significant difference.

B) A correlation heatmap illustrating the correlations among the six mentioned genes. The color of each cell in the heatmap represents the strength of correlation between two genes: blue for a negative correlation, red for a positive correlation, with deeper colors indicating stronger correlations. "*" and "**" denote the statistical significance of the correlations. For example,
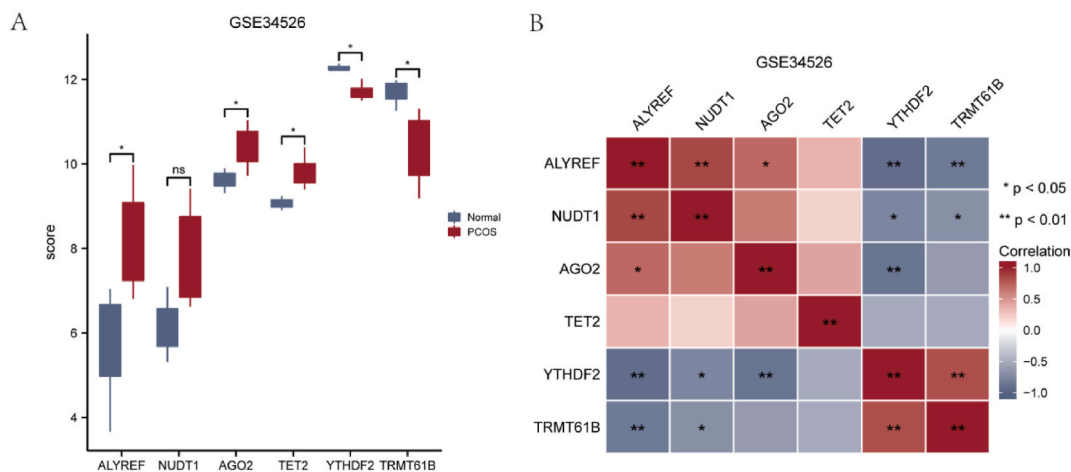
Fig. 7. Differential expression and correlation analyses of hub genes in the GSE34526 dataset
A. Results of the differential expression analysis of the GSE34526 dataset. B. The correlation between hub genes was analyzed in the GSE34526 dataset, and a heat map was drawn. ns indicates P > 0.05 (no statistical significance). * indicates P < 0.05 (moderate statistical significance). ** indicates P < 0.01 (high statistical significance). PCOS: Polycystic ovary syndrome; RMRDEGs: RNA modification related differentially expressed genes.

there is a strong positive correlation between ALYREF and NUDT1 (indicated by a deep red square with two stars), and a strong negative correlation between YTHDF2 and TET2 (indicated by a deep blue square with two stars).

Integrating the information from Fig. 7, we can deduce that in the GSE34526 dataset, there are differences in the expression levels of these six genes between normal and PCOS samples, with significant correlations observed among some of the genes.

3.7. ROC validation of hub genes in the three data sets

ROC analysis was performed to assess the discriminatory accuracy of the six hub genes (*ALYREF, NUDT1, AGO2, TET2, YTHDF2,* and *TRMT61B*) in the GSE34526, GSE5850, and GSE98421 datasets, derived from PPI network construction in the PCOS dataset. In the GSE34526 dataset, ROC validation identified *AGO2, ALYREF,* and *TRMT61B* as hub genes with high discriminatory accuracy (AUC = 0.952) between the normal and PCOS groups (Fig. 8A–D). NUDT1 showed moderate accuracy (AUC = 0.857) (Fig. 8C). The performance of TET2 and YTHDF2 is displayed in Fig. S2. In the GSE5850 dataset, ALYREF, NUDT1, and TET2 displayed low discriminatory accuracy (AUC = 0.639, 0.611, and 0.694, respectively) (Fig. 8E–G). TRMT61B and YTHDF2 demonstrated moderate accuracy (AUC = 0.750 and 0.889, respectively) (Fig. 8H and I). AGO2 is shown in Fig. S1. In the GSE98421 dataset, TRMT61B and YTHDF2 exhibited low accuracy (AUC = 0.625) (Fig. 8J and K). AGO2, ALYREF, NUDT1, and TET2 are displayed in Fig. S2.

The AUC value and confidence interval are labeled below each chart in Fig. 8. An AUC value close to 1.0 indicates high diagnostic efficacy of the gene, whereas a value closer to 0.5 suggests diagnostic performance akin to random chance. For example, in Fig. 8A, the *AGO2* gene in the GSE34526 dataset exhibits high diagnostic efficacy, with an AUC of 0.952. In Fig. 8I, the *YTHDF2* gene in the GSE5850 dataset also shows high diagnostic efficacy, with an AUC of 0.694. Conversely, in Fig. 8J, the *TRMT61B* gene in the GSE98421 dataset has lower diagnostic efficacy, with an AUC of 0.625. These ROC curves allow researchers to assess the potential of different genes as diagnostic biomarkers for PCOS.

4. Discussion

PCOS is a complex endocrine and metabolic disorder whose etiology is believed to involve both epigenetic and developmental changes influenced by hormonal imbalances in the maternal womb environment [30,31]. Epigenetic modifications, including DNA methylation, histone modification, genomic imprinting, and long non-coding RNA (lncRNA) expression, play an significant role in the pathogenesis of PCOS [32]. lncRNAs are expressed in various tissues and serve as key regulatory factors in several biological processes [33]. Differentially expressed miRNAs in the placental tissues of PCOS patients regulate target genes, contributing to the pathogenesis of PCOS [34]. Furthermore, PCOS is associated with amyotrophic lateral sclerosis [35], a disease linked to endocannabinoid system dysregulation. Additionally, RNA transport, spliceosomes, and mRNA surveillance pathways have been implicated in PCOS, influencing granulosa cell metabolism and androgen receptor regulation [35,36]. Therefore, the current study investigated differentially expressed miRNAs and their target genes, providing valuable insights into the pathogenesis of PCOS.

We constructed a PPI network for these DEGs and screened them for top-scoring gene modules. To further determine the correlation between the six hub genes, we constructed a heatmap of their correlations in the GSE34526 dataset (Fig. 8B). The results revealed highly statistically significant differences between ALYREF and NUDT1, as well as significant differences between ALYREF and
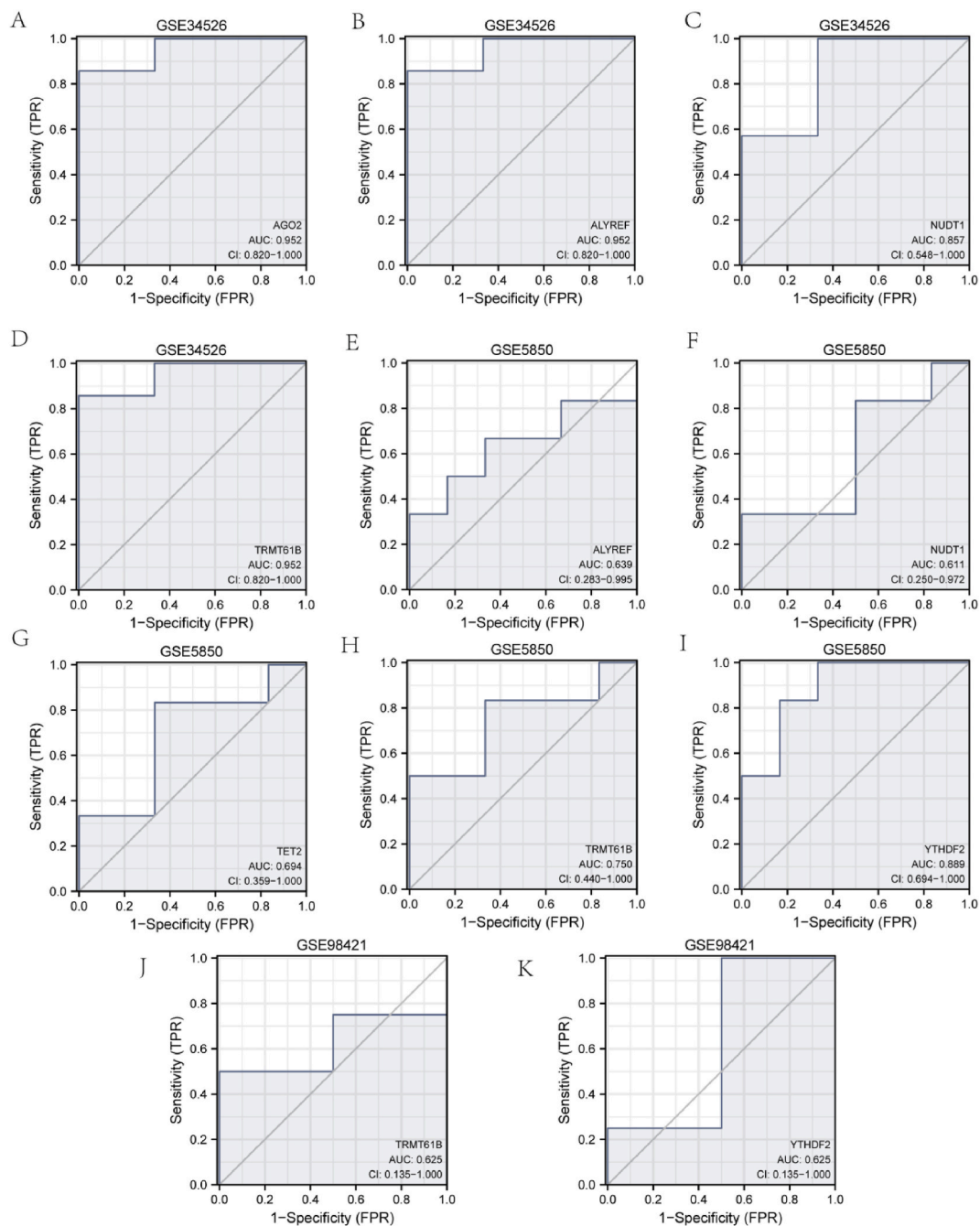
Fig. 8. ROC curve analysis of differentially expressed genes

A–D. ROC curve analysis of the DEGs *AGO2* (A), *ALYREF* (B), *NUDT1* (C), and *TRMT61B* (D) in the GSE34526 data set. E–I. ROC curve analysis of the DEGs *ALYREF* (E), *NUDT1* (F), *TET2* (G), *TRMT61B* (H), and *YTHDF2* (I) in the GSE5850 dataset. ROC curve analysis of the DEGs *TRMT61B* (J) and *YTHDF2* (K) in the GSE98421 dataset. ROC: Receiver operating characteristic; DEG: Differentially expressed gene. The closer the AUC is to 1, the greater the diagnostic effect. An AUC between 0.5 and 0.7 indicates low accuracy, an AUC between 0.7 and 0.9 indicates certain accuracy, and an AUC >0.9 indicates high accuracy.

YTHDF2. Previous studies have reported ALYREF expression in ovarian cancer. Further, AGO2 dysfunction plays a significant role in obesity in PCOS patients, and TET2 may indicate epigenetic alterations in the cumulus granulosa cells of women with PCOS. Conversely, YTHDF2 is believed to suppress tumor growth through immune reprogramming to enhance CD8$^+$ T cell-mediated anti-tumor immunity. Studies on YTHDF2 have shown that several traditional Chinese medicines can regulate m6A modification and inhibit the Wnt/β-catenin signaling pathway in rats with PCOS, promoting the recovery of ovarian morphology. Trmt61B is a

methyltransferase responsible for introducing 1-methyladenosine at position 58 of human mitochondrial tRNA. In an infertility model of PCOS, pituitary p62 mRNA was found to be consistently positively correlated with LH levels via mitochondrial OXPHOS. Therefore, these hub genes may have diagnostic and therapeutic implications for PCOS.

We constructed a PPI network for these DEGs and screened them for top-scoring gene modules. To further determine the correlation between the six hub genes, we constructed a heat map of their correlations in the GSE34526 dataset (Fig. 8B). The results revealed highly statistically significant difference between *ALYREF* and *NUDT1*, as well as significant difference between *ALYREF* and *YTHDF2*. Previous studies have reported ALYREF expression in ovarian cancer [37,38]. Further, AGO2 dysfunction plays a significant role in obesity in PCOS patients [39,40], and TET2 may indicate epigenetic alterations in the cumulus granulosa cells of women with PCOS [41]. Conversely, YTHDF2 is believed to suppress tumor growth through immune reprogramming to enhance CD8[+] T cell-mediated antitumor immunity [42]. Studies on YTHDF2 have shown that several traditional Chinese medicines can regulate m6A modification and inhibit the Wnt/β-catenin signaling pathway in rats with PCOS, promoting the recovery of ovarian morphology [43]. Trmt61B is a methyltransferase responsible for introducing 1-methyladenosine at position 58 of human mitochondrial tRNA [44]. In an infertility model of PCOS, pituitary p62 mRNA was found to be consistently positively correlated with LH levels via mitochondrial OXPHOS [45]. Therefore, these hub genes may have diagnostic and therapeutic implications for PCOS.

Pathway enrichment analyses indicated the involvement of various pathways, including the Notch, Wnt, IL2, and IL6 signaling pathways. Recent studies have demonstrated that Notch signaling plays a significant role in the ovary [41,46,47], while Wnt signaling has been linked to the effects of specific treatments for PCOS [43]. Further, a clinical study found that letrozole activated the Wnt/B-catenin pathway, leading to full endometrial proliferation in women with PCOS receiving letrozole [48]. Additionally, IL-6 signaling is associated with inflammation in ovarian syndrome [49].

In summary, while many web-based tools can directly query, retrieve, and perform differential gene expression analysis and its downstream pathway and enrichment analysis, our study is based on a combined analysis using multiple independent PCOS-related datasets (GSE34526, GSE5850, and GSE98421). Thus, the reliability and universality of the results are improved. Although differential gene expression analysis itself is not new, we specifically focused on RNA modification-associated differentially expressed genes (RMRDEGs), a relatively rare perspective in PCOS studies, which could reveal the potential role of RNA modification in PCOS pathogenesis. Based on this, we conducted in-depth GO and KEGG enrichment analyses for RMRDEGs, as well as GSEA and GSVA analyses, to gain a more detailed understanding of the functions and pathways of these genes in biological processes. Notably, we identified six hub genes (ALYREF, NUDT1, AGO2, TET2, YTHDF2, TRMT61B), among which ALYREF, NUDT1, and TRMT61B were reported for the first time in PCOS-related studies, marking an innovative aspect of this study. Additionally, we used the CIBERSORTx algorithm to predict the infiltration of different immune cell types in PCOS samples and found a significant correlation between RMRDEGs and a variety of immune cells, providing a new perspective to explore the immunological background and inflammatory mechanisms of PCOS. We constructed protein-protein interaction networks (PPI network), hub gene interaction networks with miRNA, RBP, and transcription factors, further revealing the molecular network that plays a key role in PCOS, providing valuable clues for subsequent studies. We also performed expression analysis and ROC validation of these key genes in multiple datasets, further validating their potential diagnostic value in PCOS studies.

Although this study reveals the potential pathological mechanisms of PCOS through high-throughput gene expression data and explores its molecular characteristics using bioinformatics analysis, we must acknowledge several limitations. First, the data we used are mainly derived from public databases (such as GSE34526, GSE5850, and GSE98421). These datasets typically lack detailed clinical information (such as sex hormone levels and metabolic indicators), which limits our ability to compare different PCOS subtypes and may affect the general applicability of the results. Additionally, due to variations in the sources and experimental methods of the public databases, compatibility and consistency issues may arise despite reasonable standardization, potentially leading to certain biases in the results. Secondly, GSE34526, used as the primary training set, has a small sample size, which may limit the generalizability of the conclusions. Although we validated our findings with independent datasets, the quantity and diversity of the sample population are insufficient to fully represent all PCOS patients. Future studies need to include more diverse sample populations to enhance the generalizability and stability of the conclusions. Third, the results of this study are mainly based on gene expression data from public databases. We have not yet validated the hub genes (such as ALYREF, NUDT1, AGO2, TET2, YTHDF2, TRMT61B) identified through experimental methods (such as qRT-PCR or Western blot), which limits the reliability and general applicability of the conclusions. Although validation using external datasets can enhance the persuasiveness of the results, these findings may still have certain limitations due to sample heterogeneity and differences in technical platforms.Finally, this study primarily relies on bioinformatics analysis to propose research hypotheses. Although this approach can provide valuable insights, the specific biological mechanisms still need to be confirmed through experimental validation. Future research should focus on collecting and analyzing more detailed clinical data, combining multiple data sources and experimental methods to comprehensively and thoroughly validate our findings, thereby enhancing the reliability and general applicability of the results.

In conclusion, the results of this study are consistent with many previous studies that proposed PCOS-related differences at the gene and miRNA levels. Additionally, this study confirmed the existence of immune-related differences. We suggest that these prominent differential genes (ALYREF, NUDT1, AGO2, TET2, YTHDF2, and TRMT61B) could be used as biomarkers for the diagnosis and treatment of PCOS.

Data availability statement

Data are available in a public, open access repository. Data are available on reasonable request. All data relevant to the study are included in the article or uploaded as supplemental information.The datasets (GEO data) for this study canbe found in the GEO

(https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE3452,https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE5850 and https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE98421).

Ethics approval

Not applicable.

CRediT authorship contribution statement

Kewei Quan: Writing – review & editing, Writing – original draft, Formal analysis, Data curation, Conceptualization. **Shuting Ning:** Software, Investigation. **Zilin You:** Investigation, Data curation. **Gaopi Deng:** Supervision.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:Kewei Quan reports writing assistance was provided by The 2021 Scientific Research of Guangdong Provincial Bureau of Traditional Chinese Medicine and the 2021 Basic and Applied Basic Research Fund Enterprise Joint Fund of Guangdong Province that includes: funding grants. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e36706.

References

[1] N. Ajmal, S.Z. Khan, R. Shaikh, Polycystic ovary syndrome (PCOS) and genetic predisposition: a review article, Eur. J. Obstet. Gynecol. Reprod. Biol. X (3) (2019) 100060, https://doi.org/10.1016/j.eurox.2019.100060.

[2] M.J. Khan, A. Ullah, S. Basit, Genetic basis of polycystic ovary syndrome (PCOS): current perspectives, Appl. Clin. Genet. 12 (2019) 249–260, https://doi.org/10.2147/TACG.S200341.

[3] C. Kshetrimayum, A. Sharma, V.V. Mishra, S. Kumar, Polycystic ovarian syndrome: environmental/occupational, lifestyle factors; an overview, J. Turk. Ger. Gynecol. Assoc. 20 (2019) 255–263, https://doi.org/10.4274/jtgga.galenos.2019.2018.0142. Epub.

[4] C. Wang, W. Wu, H. Yang, Z. Ye, Y. Zhao, J. Liu, L. Mu, Mendelian randomization analyses for PCOS: evidence, opportunities, and challenges, Trends Genet. 38 (2022) 468–482, https://doi.org/10.1016/j.tig.2022.01.005.

[5] M.R. Schaefer, The regulation of RNA modification systems: the next frontier in epitranscriptomics? Genes 26 (12) (2021) 345, https://doi.org/10.3390/genes12030345.

[6] C.F. Concha, P.T. Sir, S.E. Recabarren, B.F. Pérez, Epigenética del síndrome de ovario poliquístico [Epigenetics of polycystic ovary syndrome], Rev. Med. Chile 145 (2017) 907–915, https://doi.org/10.4067/s0034-98872017000700907. Spanish.

[7] Z. Chen, H. Ou, H. Wu, P. Wu, Z. Mo, Role of microRNA in the pathogenesis of polycystic ovary syndrome, DNA Cell Biol. 38 (2019) 754–762, https://doi.org/10.1089/dna.2019.4622.

[8] X. Sun, J. Lu, H. Li, B. Huang, The role of m6A on female reproduction and fertility: from gonad development to ovarian aging, Front. Cell Dev. Biol. 10 (2022) 884295, https://doi.org/10.3389/fcell.2022.884295.

[9] D. Hiam, D. Simar, R. Laker, A. Altıntaş, M. Gibson-Helm, E. Fletcher, A. Moreno-Asso, A.J. Trewin, R. Barres, N.K. Stepto, Epigenetic reprogramming of immune cells in women with PCOS impact genes controlling reproductive function, J. Clin. Endocrinol. Metab. 104 (2019) 6155–6170, https://doi.org/10.1210/jc.2019-01015.

[10] T. Barrett, et al., NCBI GEO: archive for functional genomics data sets–update, NAR 41 (2013) D991–D995, https://doi.org/10.1093/nar/gks1193.

[11] S. Chodary Khameneh, S. Razi, S. Shamdani, G. Uzan, S. Naserian, Weighted correlation network analysis revealed novel long non-coding RNAs for colorectal cancer, Sci. Rep. 12 (2022) 2990, https://doi.org/10.1038/s41598-022-06934-w.

[12] S. Kaur, K.J. Archer, M.G. Devi, A. Kriplani, JF 3rd Strauss, R. Singh, 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. Metabol. 97 (2012) E2016–E2021, https://doi.org/10.1210/jc.2011-3441.

[13] J.R. Wood, D.A. Dumesic, D.H. Abbott, J.F. Strauss, 3rd. (2007) Molecular abnormalities in oocytes from women with polycystic ovary syndrome revealed by microarray analysis, J. Clin. Endocrinol. Metabol. 92 (2007) 705–713, https://doi.org/10.1210/jc.2006-2123.

[14] Li Z. Li, B. Huang, W. Yi, F. Wang, S. Wei, H. Yan, P. Qin, D. Zou, R. Wei, N. Chen, Identification of potential early diagnostic biomarkers of sepsis, J. Inflamm. Res. 14 (2021) 621–631, https://doi.org/10.2147/jir.S298604.

[15] M. Zhao, S. Shen, C. Xue, A novel m1A-score model correlated with the immune microenvironment predicts prognosis in hepatocellular carcinoma, Front. Immunol. 13 (2022) 805967, https://doi.org/10.3389/fimmu.2022.805967.

[16] Y. Wu, D. Jiang, H. Zhang, F. Yin, P. Guo, X. Zhang, C. Bian, C. Chen, S. Li, Y. Yin, D. Böckler, J. Zhang, Y. Han, N1-Methyladenosine (m1A) regulation associated with the pathogenesis of abdominal aortic aneurysm through YTHDF3 modulating macrophage polarization, Front Cardiovasc Med 9 (2022) 883155, https://doi.org/10.3389/fcvm.2022.883155.

[17] G. Bao, T. Li, X. Guan, Y. Yao, J. Liang, Y. Xiang, X. Zhong, Comprehensive analysis of the function, immune profiles, and clinical implication of m1A regulators in lung adenocarcinoma, Front. Oncol. 12 (2022) 882292, https://doi.org/10.3389/fonc.2022.882292.

[18] M.E. Ritchie, B. Phipson, D. Wu, Y. Hu, C.W. Law, W. Shi, G.K. Smyth, Limma powers differential expression analyses for RNA-sequencing and microarray studies, Nucleic acids research 43 (2015) e47, https://doi.org/10.1093/nar/gkv007.

[19] Gene Ontology Consortium: Going Forward vol. 43, NAR, 2015, pp. D1049–D1056, https://doi.org/10.1093/nar/gku1179.

[20] M. Kanehisa, S. Goto, KEGG: kyoto encyclopedia of genes and genomes, Nucleic acids research 28 (2000) 27–30, https://doi.org/10.1093/nar/28.1.27.

[21] G. Yu, L.G. Wang, Y. Han, Q.Y. He, clusterProfiler: an R package for comparing biological themes among gene clusters, OMICS A J. Integr. Biol. 16 (2012) 284–287, https://doi.org/10.1089/omi.2011.0118.

[22] A. Subramanian, P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert, M.A. Gillette, A. Paulovich, S.L. Pomeroy, T.R. Golub, E.S. Lander, J.P. Mesirov, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, PNAS of the USA 102 (2005) 15545–15550, https://doi.org/10.1073/pnas.0506580102.

[23] S. Hänzelmann, R. Castelo, J. Guinney, GSVA: gene set variation analysis for microarray and RNA-seq data, BMC Bioinf. 14 (2013) 7, https://doi.org/10.1186/1471-2105-14-7.

[24] C.B. Steen, C.L. Liu, A.A. Alizadeh, A.M. Newman, Profiling cell type abundance and expression in bulk tissues with CIBERSORTx, Methods Mol. Biol. 2117 (2020) 135–157, https://doi.org/10.1007/978-1-0716-0301-7_7.

[25] B. Wakonig, A.M.I. Auersperg, M. O'Hara, String-pulling in the Goffin's cockatoo (Cacatua goffiniana), Learn. Behav. 49 (2021) 124–136, https://doi.org/10.3758/s13420-020-00454-1.

[26] P. Shannon, et al., Cytoscape: a software environment for integrated models of biomolecular interaction networks, Genome Res. 13 (2003) 2498–2504, https://doi.org/10.1101/gr.1239303.

[27] W. Wang, J. Zhang, Y. Wang, Y. Xu, S. Zhang, Identifies microtubule-binding protein CSPP1 as a novel cancer biomarker associated with ferroptosis and tumor microenvironment, Compt Struct Biotechnol 20 (2022) 3322–3335, https://doi.org/10.1016/j.csbj.2022.06.046.

[28] Y. Chen, X. Wang, miRDB: an online database for prediction of functional microRNA targets, NAR 48 (2020) D127–d131, https://doi.org/10.1093/nar/gkz757.

[29] K.R. Zhou, S. Liu, W.J. Sun, L.L. Zheng, H. Zhou, J.H. Yang, L.H. Qu, ChIPBase v2.0: decoding transcriptional regulatory networks of non-coding RNAs and protein-coding genes from ChIP-seq data, NAR 45 (2017) D43–d50, https://doi.org/10.1093/nar/gkw965.

[30] S. Risal, Y. Pei, H. Lu, M. Manti, R. Fornes, H.P. Pui, Z. Zhao, J. Massart, C. Ohlsson, E. Lindgren, N. Crisosto, M. Maliqueo, B. Echiburú, A. Ladrón de Guevara, T. Sir-Petermann, H. Larsson, M.A. Rosenqvist, C.E. Cesta, A. Benrick, Q. Deng, E. Stener-Victorin, Prenatal androgen exposure and transgenerational susceptibility to polycystic ovary syndrome, Nat. Med. 25 (2019) 1894–1904, https://doi.org/10.1038/s41591-019-0666-1.

[31] B. Tata, N.E.H. Mimouni, A.L. Barbotin, S.A. Malone, A. Loyens, P. Pigny, D. Dewailly, S. Catteau-Jonard, I. Sundström-Poromaa, T.T. Piltonen, F. Dal Bello, C. Medana, V. Prevot, J. Clasadonte, P. Giacobini, Elevated prenatal anti-Müllerian hormone reprograms the fetus and induces polycystic ovary syndrome in adulthood, Nat. Med. 24 (2018) 834–846, https://doi.org/10.1038/s41591-018-0035-5.

[32] E. Stener-Victorin, Q. Deng, Epigenetic inheritance of polycystic ovary syndrome - challenges and opportunities for treatment, Nat. Rev. Endocrinol. 17 (2021) 521–533, https://doi.org/10.1038/s41574-021-00517-x.

[33] L. Mu, X. Sun, M. Tu, D. Zhang, Non-coding RNAs in polycystic ovary syndrome: a systematic review and meta-analysis, Reprod. Biol. Endocrinol. 19 (2021) 10, https://doi.org/10.1186/s12958-020-00687-9.

[34] O.S. Walker, A.C. Holloway, S. Raha, The role of the endocannabinoid system in female reproductive tissues, J. Ovarian Res. 12 (2019) 3, https://doi.org/10.1186/s13048-018-0478-9.

[35] X. Liu, C. Sun, K. Zou, C. Li, X. Chen, H. Gu, Z. Zhou, Z. Yang, Y. Tu, N. Qin, Y. Zhao, Y. Wu, Y. Meng, G. Ding, X. Liu, J. Sheng, C. Yu, H. Huang, Novel PGK1 determines SKP2-dependent AR stability and reprograms granular cell glucose metabolism facilitating ovulation dysfunction, EBioMedicine 61 (2020) 103058, https://doi.org/10.1016/j.ebiom.2020.103058.

[36] A.G. Matera, Z. Wang, A day in the life of the spliceosome, Nat. Rev. Mol. Cell Biol. 15 (2014) 108–121, https://doi.org/10.1038/nrm3742.

[37] P. Zheng, N. Li, X. Zhan, Ovarian cancer subtypes based on the regulatory genes of RNA modifications: novel prediction model of prognosis, Front. Endocrinol. 13 (2022) 972341, https://doi.org/10.3389/fendo.2022.972341.

[38] R. Kaaks, R.T. Fortner, A. Hüsing, M. Barrdahl, M. Hopper, T. Johnson, et al., Tumor-associated autoantibodies as early detection markers for ovarian cancer? A prospective evaluation, Int. J. Cancer 143 (2018) 515–526, https://doi.org/10.1002/ijc.31335.

[39] L. Qin, J. Chen, L. Tang, T. Zuo, H. Chen, R. Gao, W. Xu, Significant role of dicer and miR-223 in adipose tissue of polycystic ovary syndrome patients, BioMed Res. Int. 2019 (2019) 9193236, https://doi.org/10.1155/2019/9193236.

[40] L. Deng, Q. Chen, J. Xie, W. Wei, H. Hui, circPUM1 promotes polycystic ovary syndrome progression by sponging to miR-760, Gene 754 (2020) 144903, https://doi.org/10.1016/j.gene.2020.144903.

[41] P. Sagvekar, G. Shinde, V. Mangoli, S.K. Desai, S. Mukherjee, Evidence for TET-mediated DNA demethylation as an epigenetic alteration in cumulus granulosa cells of women with polycystic ovary syndrome, Mol. Hum. Reprod. 28 (2022) gaac019, https://doi.org/10.1093/molehr/gaac019.

[42] S. Ma, B. Sun, S. Duan, J. Han, T. Barr, J. Zhang, M.B. Bissonnette, M. Kortylewski, C. He, J. Chen, M.A. Caligiuri, J. Yu, YTHDF2 orchestrates tumor-associated macrophage reprogramming and controls antitumor immunity through CD8+ T cells, Nat. Immunol. 24 (2023) 255–266, https://doi.org/10.1038/s41590-022-01398-6.

[43] Y. Zhang, H. Zhou, C. Ding, The ameliorative effect of CangFu Daotan Decoction on polycystic ovary syndrome of rodent model is associated with m6A methylation and Wnt/β-catenin pathway, Gynecol. Endocrinol. 39 (2023) 2181637, https://doi.org/10.1080/09513590.2023.2181637.

[44] T. Chujo, T. Suzuki, Trmt61B is a methyltransferase responsible for 1-methyladenosine at position 58 of human mitochondrial tRNAs, RNA 18 (2012) 2269–2276, https://doi.org/10.1261/rna.035600.112.

[45] X. Li, L. Zhou, G. Peng, M. Liao, L. Zhang, H. Hu, L. Long, X. Tang, H. Qu, J. Shao, H. Zheng, M. Long, Pituitary P62 deficiency leads to female infertility by impairing luteinizing hormone production, Exp. Mol. Med. 53 (2021) 1238–1249, https://doi.org/10.1038/s12276-021-00661-4.

[46] H. Koike, H. Harada, A. Kusamoto, C. Kunitomi, Z. Xu, T. Tanaka, Y. Urata, E. Nose, N. Takahashi, O. Wada-Hiraike, Y. Hirota, K. Koga, Y. Osuga, Notch signaling induced by endoplasmic reticulum stress regulates cumulus-oocyte complex expansion in polycystic ovary syndrome, Biomolecules 12 (2022) 1037, https://doi.org/10.3390/biom12081037.

[47] Q. Han, W. Zhang, J. Meng, L. Ma, A. Li, LncRNA-LET inhibits cell viability, migration and EMT while induces apoptosis by up-regulation of TIMP2 in human granulosa-like tumor cell line KGN, Biomed. Pharmacother. 100 (2018) 250–256, https://doi.org/10.1016/j.biopha.2018.01.162.

[48] S. Mehdinejadiani, F. Amidi, M. Mehdizadeh, M. Barati, A. Pazhohan, A. Alyasin, K. Mehdinejadiani, A. Sobhani, Effects of letrozole and clomiphene citrate on Wnt signaling pathway in endometrium of polycystic ovarian syndrome and healthy women, Biol. Reprod. 100 (2019) 641–648, https://doi.org/10.1093/biolre/ioy187.

[49] A. Borthakur, Y. D Prabhu, A. Valsala Gopalakrishnan, Role of IL-6 signalling in polycystic ovarian syndrome associated inflammation, J. Reprod. Immunol. 141 (2020) 103155, https://doi.org/10.1016/j.jri.2020.103155.