Sikiru *et al.*
Middle East Fertility Society Journal (2025) 30:28
https://doi.org/10.1186/s43043-025-00241-w

Middle East Fertility Society Journal

RESEARCH

Open Access

Transcriptomic profiling of cumulus cells reveals dysregulated genes and pathways in PCOS-related infertility

Akeem Babatunde Sikiru[1,2*], Nurulfiza Mat Isa[2], Kasim Sakran Abass[3], Muibat Adesola Adeniran[4], Stephen Sunday Acheneje Egena[5] and Karimot Akinola[6]

Abstract

Background Polycystic ovarian syndrome (PCOS) is a leading cause of infertility and metabolic dysfunction in women, characterized by hyperandrogenism, anovulation, and insulin resistance. Cumulus cells play a crucial role in folliculogenesis and oocyte maturation, necessitating a deeper understanding of their molecular alterations impact in PCOS.

Method This study investigates transcriptomic differences in cumulus cells between PCOS and non-PCOS women using high-throughput RNA sequencing data obtained from the NCBI Gene Expression Omnibus (GEO) database (accession number: GSE277906). The RNA sequencing data from 23 PCOS and 17 non-PCOS women were analyzed to identify differentially expressed genes (DEGs) using R-based computational pipelines.

Results Differential gene expression analysis identified 3245 significantly dysregulated genes, comprising 1723 upregulated and 1522 downregulated genes in PCOS samples. Functional enrichment analysis revealed that key DEGs (CDH5, CLEC4D, and GNAT1) were associated with follicular development, insulin signaling, and immune response. Gene Set Enrichment Analysis (GSEA) further identified dysregulation in metabolic and reproductive pathways, including ribonucleoprotein complex biogenesis and vascular endothelial growth factor (VEGF) signaling.

Conclusion This study highlights that altered gene expression in cumulus cells may impair oocyte competence, potentially influencing fertility outcomes in PCOS patients. GNAT1, previously linked to diabetes, emerged as a novel gene potentially involved in PCOS pathophysiology. However, these findings are derived from a single-center dataset which requires experimental validation. Future studies should incorporate qRT-PCR validation and functional assays in larger and ethically diverse cohorts as means for development of targeted therapeutic interventions to mitigate the reproductive consequences of PCOS.

Keywords PCOS, Cumulus cells, GNAT1, Oocyte competence, Infertility

*Correspondence:
Akeem Babatunde Sikiru
akeem@upm.edu.my; akeembaba01@gmail.com
Full list of author information is available at the end of the article

Background

Polycystic ovarian syndrome (PCOS) is one of the most prevalent endocrine disorders affecting reproductive-aged women, with an estimated global prevalence of 6–20% [1, 2]. It is characterized by anovulation, hyperandrogenism, and polycystic ovarian morphology, frequently accompanied by metabolic conditions such as insulin resistance and type 2 diabetes mellitus [3, 4]. The complexity of PCOS extends beyond reproductive dysfunction, affecting systemic metabolism, inflammation, and cardiovascular health, making its management particularly challenging given the fact that it has a strong genetic complexity [2].

Cumulus cells, which surround the oocyte, are essential for folliculogenesis, oocyte maturation, and fertilization [5]. They play a vital role in metabolic support and the regulation of local endocrine signaling, which in turn affects oocyte quality and overall reproductive success [6, 7]. Although, several studies have explored the transcriptomic profiles of granulosa cells in relation to PCOS leading to identification of dysregulated pathways including steroidogenesis, oxidative stress and immune signaling. However, there are reviews and other nonempirical data establishing cumulus cells as critical components of ovarian physiology, and gene-level expression changes and functional implications remain underexplored. Meanwhile, understanding how the disruptions in pattern of gene expression within cumulus cells may impair follicular development and poor oocyte competence, affecting fertility outcomes in women with PCOS [2].

Therefore, this study fills knowledge gaps using publicly available RNA-seq dataset (GSE277906) and applying comprehensive computational analysis comparing gene expression profiles in cumulus cells from PCOS and non-PCOS women. By identifying differentially expressed genes (DEGs) and enriched biological pathways, this study aims to elucidate the molecular mechanisms underlying PCOS pathophysiology. This was undertaken because understanding these transcriptional changes could pave the way for novel biomarkers and therapeutic strategies aimed at improving ovarian function and fertility outcomes in PCOS patients. This study investigates differences in mRNA expression in cumulus cells between women with polycystic ovary syndrome (PCOS) and non-PCOS women.

Methods

Source data and sequencing overview

The dataset used was derived from GSE277906, consisting of cumulus cell samples from 23 PCOS and 17 non-PCOS women undergoing fertility treatment at a clinic in China. The samples were sequenced using the Illumina NovaSeq 6000 platform. Demographic and clinical metadata such as BMI, hormonal profiles, and ovulatory status were not available publicly, which precluded phenotype-specific analysis and is acknowledged as a limitation [8]. The publicly available data were analyzed to elucidate metabolic alterations in follicular environments that could influence reproductive function in PCOS patients.

The clinical data associated with the dataset from the original study reported no significant differences between PCOS and control groups in age, body mass index (BMI), follicle-stimulating hormone (FSH), total testosterone (TT), serum total cholesterol (TC), or fasting glucose levels. However, patients with PCOS showed significantly higher levels of antral follicle count (AFC), anti-Müllerian hormone (AMH), luteinizing hormone (LH), LH/FSH ratio, and triglycerides (TG) concentrations [8].

RNA-seq data processing

The Fragments Per Kilobase of transcript per Million mapped reads (FPKM) RNA-seq data was processed using RStudio (version 2024.12.1 + 563 for Windows). The steps included data import whereby the FPKM data file was loaded using the readxl package. Data structuring was also conducted whereby the gene identifiers were set as row names and sample identifiers were used as column names. The quality control followed handling of missing values using the na.rm = TRUE argument, and overall distribution was assessed for downstream analysis [9].

Differential genes expression analysis and their statistics

Differentially expressed genes (DEGs) were identified by first programmatically categorizing the samples into "control" and "PCOS" groups. The mean expression of each gene across the 28,001 genes was calculated for both groups. Fold change (FC) and log2 fold change (log2FC) values were computed using the dplyr package in R software to determine the magnitude of expression differences [10]. A Student's *t*-test was then applied on a per-gene basis to compare expression levels between the PCOS and control groups, generating *p*-values for statistical significance. The genes were classified as upregulated ($FC \geq 1, p < 0.05$) or downregulated ($FC < 1, p < 0.05$). To account for multiple testing and reduce false positives, a false discovery rate (FDR) correction was applied using the Benjamini–Hochberg method. The classification of genes into significant and nonsignificant categories, along with the associated parameters, was utilized as a data processing strategy to enhance our understanding of the functional roles of these genes within each regulatory category [11].

Statistical analysis

Basic descriptive statistics were calculated for the 28,001 genes to provide an overall summary of gene expression patterns. These included measures such as the mean, median, and standard error of the mean (SEM) to assess central tendency and variability. Additionally, the minimum, third quartile (Q3), and maximum values were determined to capture the range and distribution of expression levels across samples. All these computations were performed using built-in R functions, including mean(), median(), quantile(), and max(), ensuring that missing values were managed appropriately by setting the na.rm argument to TRUE [12]. This statistical analysis provided a foundation for further interpretation of gene expression differences between PCOS and control samples.

Visualization of differentially and variedly expressed genes

To enhance the interpretability of the gene expression patterns, several visualization techniques were employed. Boxplots were generated using ggplot2 in R and Seaborn in Python to display the expression levels of the top 10 differentially expressed genes, providing a clear comparison between the PCOS and control groups [13, 14]. Additionally, a heatmap was created using the ComplexHeatmap package in R to visualize the expression patterns of the 50 most variable genes, highlighting clustering patterns that distinguish the 2 groups [15]. To examine the overall distribution of log2 fold change (log2FC) values, a histogram was constructed with ggplot2, offering insights into the skewness and spread of differential expression. Lastly, a volcano plot was generated to illustrate the relationship between the magnitude of gene expression changes and their statistical significance, allowing for the identification of key upregulated and downregulated genes. These visualization techniques were conducted to collectively provide a comprehensive overview of the gene expression dynamics in the PCOS group.

Gene Set Enrichment Analysis

Gene Set Enrichment Analysis (GSEA) was conducted to identify biological processes associated with the differentially expressed genes (DEGs) in PCOS using the clusterProfiler package [16]. To perform the analysis, genes were first sorted based on their log2 fold change (log2FC) values, prioritizing those with the most significant expression differences. The gseGO function was then applied to determine enriched Gene Ontology (GO) biological processes, allowing for a deeper understanding of functional pathways influenced by dysregulated genes. Additionally, pathway enrichment analysis was performed using g:Profiler [17], using the top 10 significantly differentially upregulated genes including CDH5, CLEC4D, GNAT1, GPR25, KRTAP7-1, LINC00398, LINC02269, LOC101928295, RGS7BP, and RUNX3-AS1 in the PCOS group to ensure that the most relevant pathways were highlighted. The complete list of differentially expressed genes (DEGs) was saved, including the list of the significantly upregulated genes in a plain text file (DEG_Gene_List.txt) for further analysis. The scripts used for preprocessing, statistical analysis, and visualization were written in R and Python and are available as supplementary materials.

Table 1 Summary of differential gene expression analysis in PCOS and control samples

Parameters	Value
Total number of genes analyzed	28,001
Significantly differentially expressed genes (*FDR* < 0.05)	3245
Significantly upregulated genes in PCOS	1723
Significantly downregulated genes in PCOS	1522
Average expression value (FPKM)	3.21
Median expression value (FPKM)	2.45
Standard error of the mean (SEM)	0.89

This table summarizes the key findings from the differential gene expression analysis comparing PCOS and control samples. It highlights the total number of genes, significantly differentially expressed genes, and their regulation patterns. Expression values are reported in Fragments Per Kilobase of transcript per Million mapped reads (*FPKM*). False discovery rate (*FDR*) correction was applied to determine statistical significance. *PCOS*, polycystic ovary syndrome; *FDR*, false discovery rate; *FPKM*, Fragments Per Kilobase of transcript per Million mapped reads

Results

Differential expressed genes (DEGs)

A total of 28,001 genes were analyzed to identify differences in gene expression between PCOS and control samples. Out of these, 3245 genes were found to be significantly differentially expressed based on an adjusted *p*-value threshold of 0.05 (FDR adjusted). Among these differentially expressed genes, 1723 were upregulated in PCOS samples, while 1522 genes were downregulated. The distribution of gene expression across all samples showed an average expression value of 3.21 FPKM, a median of 2.45 FPKM, and a standard error of the mean (SEM) of 0.89 FPKM (Table 1). These findings highlight significant transcriptional alterations in cumulus cells, suggesting potential molecular pathways that may be dysregulated in PCOS.

Visualization of the gene expression patterns

To further explore the differences in gene expression between PCOS and control samples, several visualization techniques were employed. Boxplots were generated to examine the expression levels of the top 10 differentially

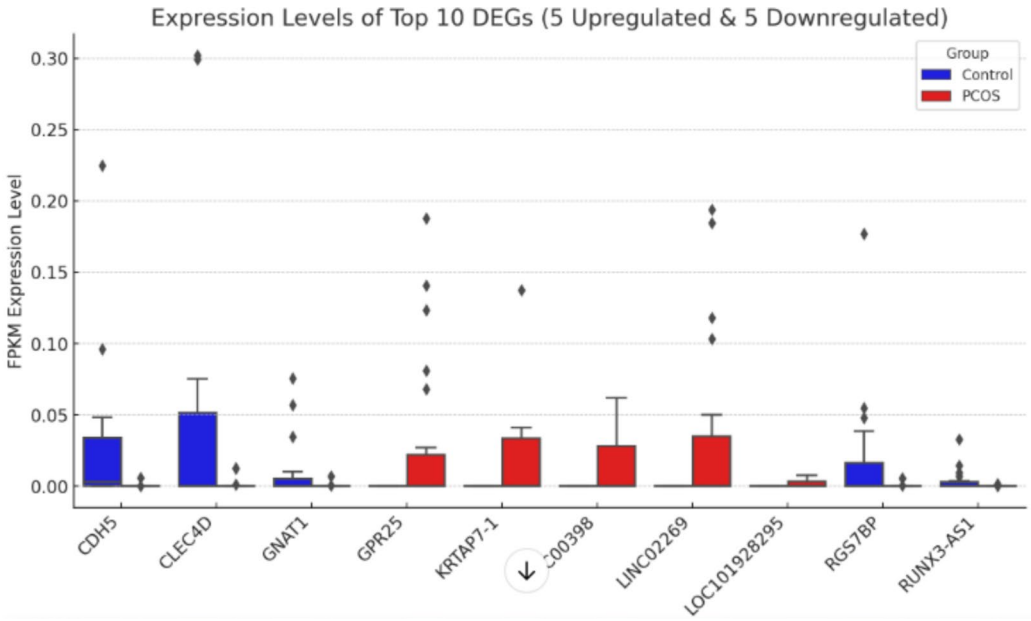

Fig. 1 Expression levels of top 10 differentially expressed genes (DEGs) in PCOS versus the control. This plot displays the expression levels (measured as FPKM — Fragments Per Kilobase of transcript per Million mapped reads) of the top 10 differentially expressed genes (DEGs) between the PCOS and control groups. The genes are ranked by their significance of differential expression, with the five most upregulated genes and the five most downregulated genes shown. The *x*-axis represents the gene names. The *y*-axis represents the FPKM expression level. The data is presented as boxplots, where each box represents the interquartile range (IQR) of the data, the line inside the box represents the median, and the whiskers extend to 1.5 times the IQR. Individual data points outside this range are shown as diamonds. The bars are colored to distinguish between the two groups: blue for the control group and red for the PCOS group. This visualization allows for a direct comparison of the expression levels of these key genes between the two groups, highlighting the magnitude and direction of differential expression

expressed genes, confirming distinct alterations in gene regulation between the 2 groups (Fig. 1). Apart from the top differentially expressed gene, the volcano plot showed an overall pattern of the gene expressions (Fig. 2). The heatmap provided a comprehensive view of clustering patterns among highly variable genes ($n=50$), demonstrating distinct gene expression profiles in the PCOS samples compared to the controls (Fig. 3). Additionally, the histogram of log2 fold change (log2FC) values revealed a skewed distribution, indicating that a subset of genes exhibited substantial expression changes (Fig. 4). These visualizations showed relationships between the magnitude of gene expression changes and their statistical significance, providing insights into the molecular alterations associated with the PCOS phenotype.

Gene set enrichment and functional pathway analysis

GSEA was performed to identify biological processes and molecular pathways enriched among the differentially expressed genes. The analysis revealed several key pathways significantly associated with PCOS. Notably, follicular development was enriched, highlighting disruptions in ovarian function that may impact oocyte quality. Additionally, the insulin signaling pathway was significantly enriched, supporting previous evidence linking PCOS to insulin resistance and metabolic dysfunction. Another important pathway, ribonucleoprotein complex biogenesis, was identified, indicating alterations in hormone production that may contribute to the pathophysiology of PCOS (Fig. 5). Further pathway analysis using g:Profiler confirmed enrichment in metabolic and reproductive pathways, reinforcing the role of dysregulated gene expression in the molecular mechanisms underlying PCOS (Table 2). These findings suggest potential targets for therapeutic intervention and provide a deeper understanding of the biological processes involved in the disorder.

Discussion

This study identified significant transcriptional changes in cumulus cells of women with PCOS; key DEGs such as CDH5, CLEC4D, and GNAT1 were enriched in pathways associated with follicular development, insulin signaling, and immune response. In addition to the original study from which the dataset was generated, this study focuses on gene-level DEGs and functional annotations, offering new insights such as the novel association of GNAT1 with PCOS. While the dataset's single-center origin may limit generalizability, the absence of clinical metadata such as age, BMI,

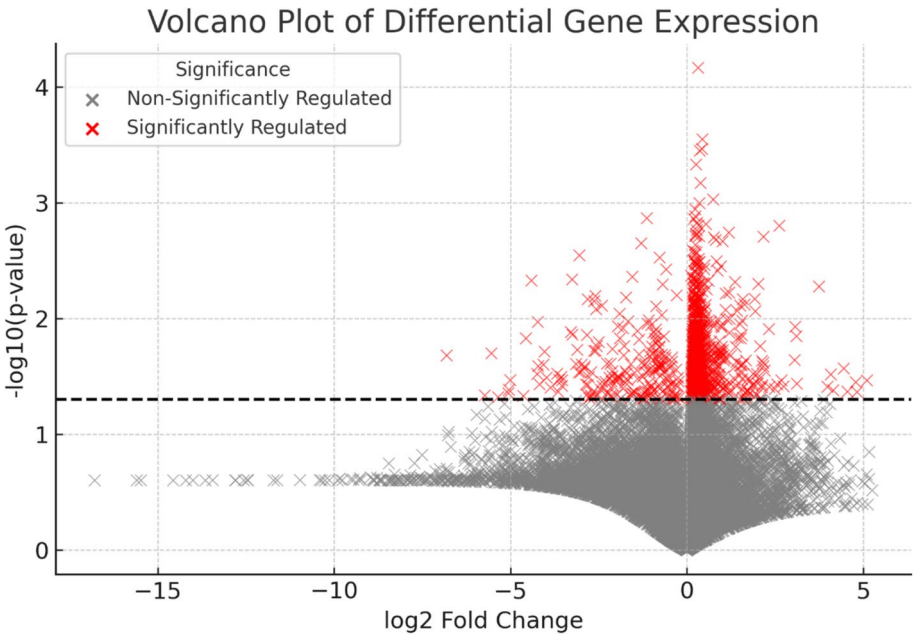

Fig. 2 Volcano plot of differential gene expression. This volcano plot displays the results of the differential gene expression analysis, showing the relationship between the magnitude of gene expression change (log2 fold change) and its statistical significance (-log10(*p*-value)). Each point represents a gene; the genes that are not significantly regulated are shown in gray color, while genes that are significantly regulated are shown in red color. The horizontal dashed line indicates the significance threshold (typically -log10(*p*-value) = 1.3, corresponding to $p = 0.05$), and the genes above this line are considered statistically significant. The *x*-axis represents the log2 fold change, where values greater than zero indicate upregulation and values less than zero indicate downregulation. The further a point is from the center along the *x*-axis, the greater the magnitude of the change in gene expression. This plot helps to visualize which genes have both a large magnitude of change and high statistical significance in expression

or hormonal status also limits stratified analysis. Furthermore, while GSEA indicates potential mechanistic pathways, causality remains unconfirmed without functional validation. Future work should include qRT-PCR validation in independent populations and in vitro studies using cumulus cell cultures to validate the biological roles of identified genes. This study thus provides a foundation for future biomarker discovery and therapeutic development.

This study identified significant transcriptional changes in cumulus cells of women with PCOS, revealing differentially expressed genes (DEGs) upregulated in the PCOS women compared to non-PCOS controls, where they were downregulated. The visualization highlighted key clusters of differentially expressed genes, reflecting distinct gene expression profiles in PCOS. Notably, CDH5, CLEC4D, and GNAT1 emerged as significant DEGs, with their involvement in processes such as ovarian follicle development, immune regulation, and hormone receptor signaling. The GSEA further revealed enrichment in biological pathways, including follicular development, insulin signaling, and ribonucleoprotein complex biogenesis. These findings suggest that the altered molecular environment in cumulus cells could impact oocyte competence, influencing fertility outcomes in PCOS patients [8].

Moreover, the interplay between these differentially expressed genes and metabolic pathways signifies the multifactorial nature of PCOS. The dysregulation of key genes involved in metabolic homeostasis and cellular signaling may contribute to the compromised microenvironment surrounding the oocyte and consequentially could cause infertility in PCOS patients [18]. Notably, the aberrant expression of genes linked to oxidative stress and inflammatory responses may further exacerbate follicular dysfunction [2]. This highlights the potential role of targeted therapeutic strategies aimed at modulating these molecular pathways to improve reproductive outcomes in PCOS-affected individuals.

These findings align with the long-standing relationship between metabolic and hormonal dysregulation and the pathogenesis of PCOS and can be linked to the enrichment of genes involved in insulin signaling, which supports the well-established association between PCOS and insulin resistance. This was reported in several other prior studies that highlighted insulin receptor (INSR) dysregulation in PCOS granulosa and cumulus cells, contributing to aberrant follicular development and

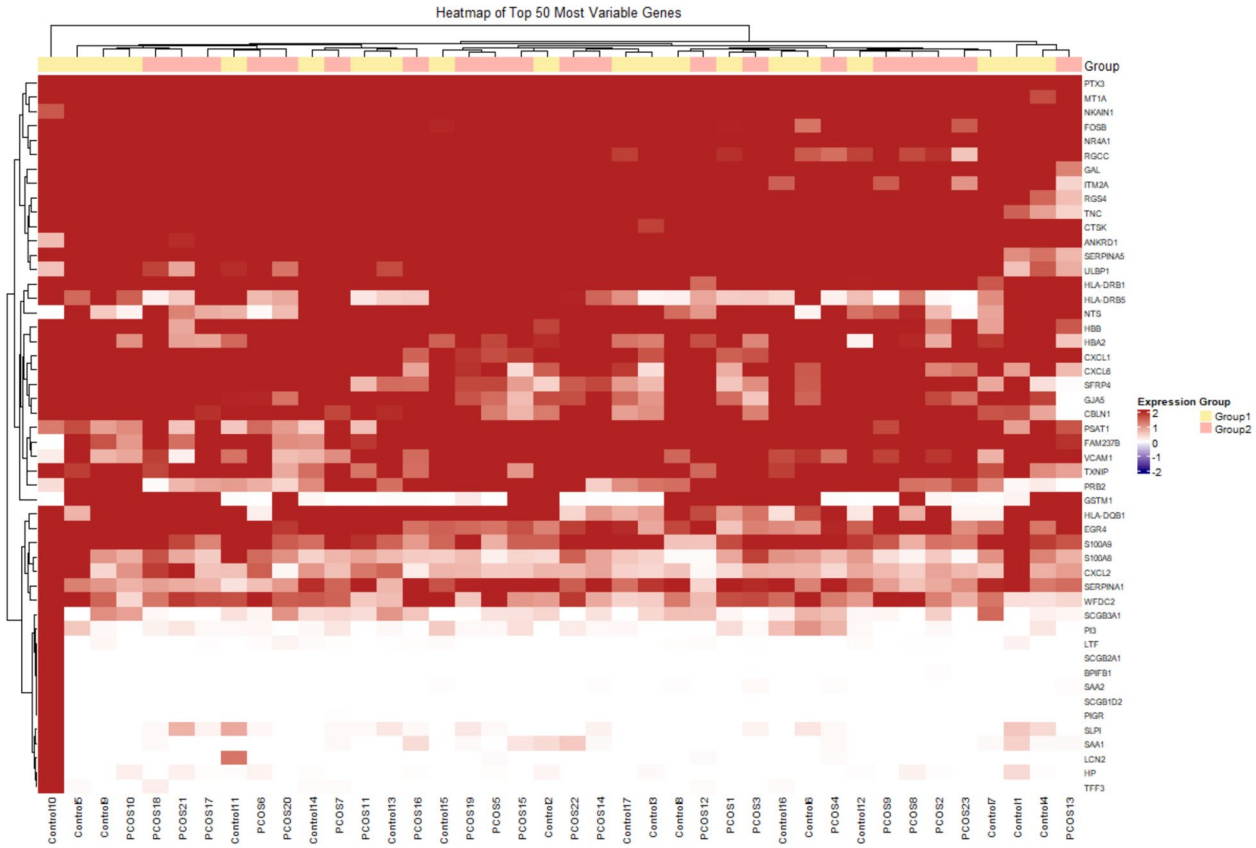

Fig. 3 Heatmap of top 50 most variable genes across samples. This heatmap visualizes the expression levels of the 50 most variable genes across different samples. The rows represent the genes, and each column represents a sample. The color gradient, indicated by the legend, represents the gene expression level, with red indicating higher expression and blue indicating lower expression. The samples are grouped into two categories as shown by the color bar at the top. The dendrogram on the left shows the hierarchical clustering of the genes based on their expression patterns, revealing groups of co-expressed genes. The sample names are displayed at the bottom of the heatmap, allowing for easy identification of individual samples. This visualization helps to identify gene expression patterns and differences between the two sample groups

hyperandrogenism [19, 20]. Similarly, CDH5, which was significantly differentially expressed in this present study, has been implicated in vascular endothelial growth factor (VEGF) signaling, a pathway known to regulate follicular angiogenesis and ovarian function [21].

The overexpression of CLEC4D, involved in immune signaling, suggests a potential inflammatory component in PCOS, consistent with existing literature indicating chronic low-grade inflammation as a characteristic feature of the syndrome [22]. Interestingly, GNAT1, a gene associated with phototransduction, was found to be significantly dysregulated in this study. While GNAT1 has not been previously linked to PCOS, emerging evidence suggests that the gene is associated with diabetes [23], which is a known comorbidity with PCOS. This could indicate a role for G-protein signaling pathways in ovarian follicle maturation, warranting further investigation. While many findings in this study are consistent with established literature, the differential regulation of specific genes such as GNAT1 introduces novel insights that merit further validation.

The altered gene expression in cumulus cells suggests disruptions in key biological processes critical for oocyte competence and follicular health [7, 24, 25]. However, the enrichment of ribonucleoprotein (RNP) complex biogenesis pathways indicates an alteration in RNA processing and protein translation in cumulus cells, which may contribute to defective oocyte maturation. The RNP has been reported to be crucial for vital cellular functions like transcription, translation, and gene regulation, impacting development, cell function, and disease development; therefore, its understanding in relation to PCOS could be critical [26]. The identification of differentially expressed genes involved in follicle-stimulating hormone receptor (FSHR) activity suggests potential dysregulation of gonadotropin signaling, a key factor in follicular development and ovulation [27, 28]. Additionally, the enrichment of genes involved in VEGF receptor

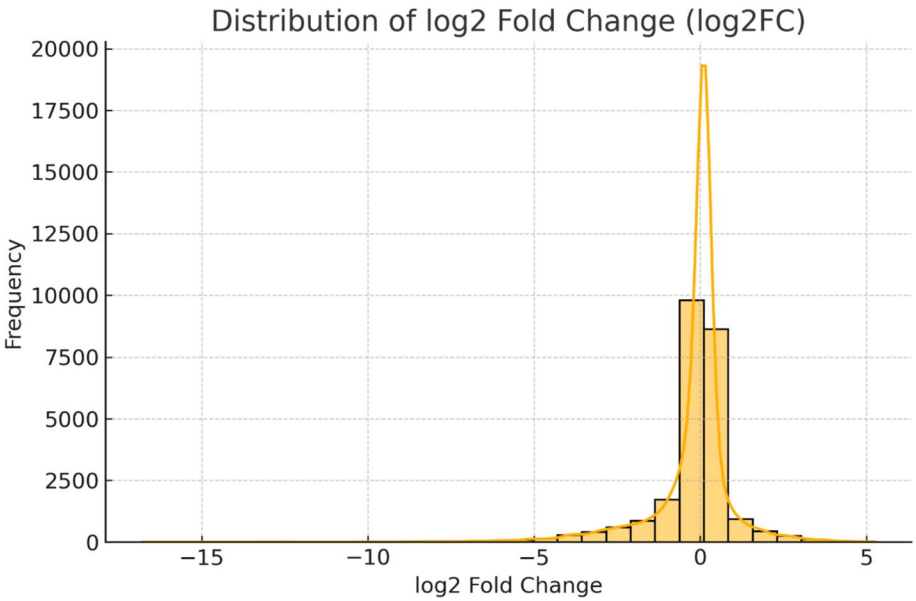

Fig. 4 Distribution of log2 fold change (log2FC). This plot displays the distribution of log2 fold change (log2FC) values observed in the differential gene expression analysis. The *x*-axis represents the log2FC, indicating the magnitude and direction of gene expression change between the control and the PCOS conditions. The log2FC value 0 indicates no change, positive values indicate upregulation, and negative values indicate downregulation. The *y*-axis represents the frequency, showing the number of genes falling within each log2FC bin. The histogram bars represent the actual frequency counts for each bin, while the overlaid curve represents a kernel density estimation, providing a smoothed representation of the data distribution. The plot highlights the overall spread and central tendency of log2FC values, revealing the extent of gene expression changes and potential biases or trends in the data

binding and endothelial cell adhesion suggests that vascular dysfunction may contribute to abnormal folliculogenesis in PCOS. Clinically, these findings could have significant implications in the use of the identified DEGs as potential biomarkers for early PCOS diagnosis or therapeutic targets for improving ovarian function in affected women. For example, modulating VEGF signaling could enhance follicular angiogenesis, potentially improving oocyte quality and fertility outcomes in PCOS patients.

Although, this study leveraged high-throughput RNA sequencing to provide an in-depth transcriptomic profile of cumulus cells in PCOS, offering novel insights into molecular mechanisms underlying the disorder. Also, the use of stringent statistical analyses including FDR correction enhances the reliability of the findings. Additionally, the study incorporated multiple visualization techniques to improve the interpretability of gene expression patterns. However, the limitations such as the derivation of the dataset from a publicly available repository may introduce batch effects and variability in sample collection or processing methods. Additionally, while gene expression data suggest potential regulatory mechanisms that are reported through the functional enrichment analysis, the validation of these pathways through in vitro or in vivo models is necessary to establish causality of the identified pathways and gene regulations. Further, the samples were obtained in a Chinese population; future studies with larger and ethnically diverse cohorts may provide a more comprehensive understanding of the transcriptomic landscape in PCOS.

There is also a need for future research that could focus on the validation of the top identified DEGs and pathways through functional studies. Experimental approaches such as CRISPR-based gene editing or siRNA knockdown studies in cumulus cell cultures could help determine the mechanistic role of key genes such as CDH5 and GNAT1 in follicular development. Additionally, integrating proteomics and metabolomics with transcriptomics findings could provide a more holistic understanding of how transcriptomic changes translate into functional protein and metabolic alterations in PCOS. Furthermore, expanding similar studies to include a broader patient cohort with varying PCOS phenotypes such as hyperandrogenic vs. normoandrogenic PCOS groups may also help elucidate phenotype-specific molecular mechanisms. Lastly, exploring therapeutic interventions targeting dysregulated pathways, such as modulating VEGF or insulin signaling, could pave the way for novel treatment strategies aimed at improving reproductive outcomes in PCOS patients.

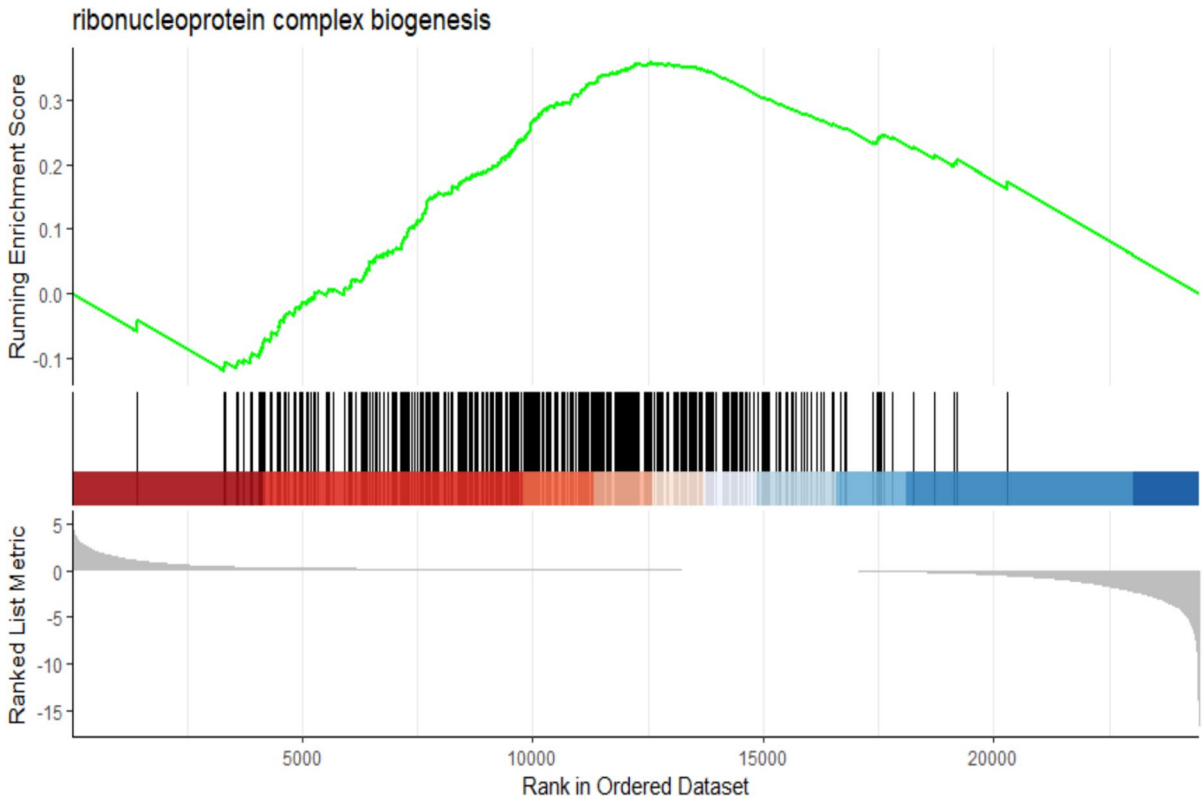

Fig. 5 The GSEA plot for ribonucleoprotein complex biogenesis in the context of polycystic ovary syndrome (PCOS). Top panel (green line): Shows the running enrichment score (ES) as the analysis walks down the ranked list of genes, ordered by their correlation with the PCOS phenotype. The ES reflects the degree to which genes in the ribonucleoprotein complex biogenesis gene set are overrepresented at the top or bottom of the ranked list. A positive ES indicates enrichment at the top of the list (genes in the set are positively correlated with PCOS), while a negative ES indicates enrichment at the bottom (genes in the set are negatively correlated with PCOS). Middle panel (black tick marks): Represents the position of the genes in the ribonucleoprotein complex biogenesis gene set within the ranked list of genes. The clustering of ticks towards the left or right indicates enrichment at the top or bottom of the list, respectively. The color gradient below the ticks indicates the correlation of the genes with the PCOS phenotype, with red representing positive correlation and blue representing negative correlation. Bottom panel (gray bars): Shows the ranked list metric, which is typically the correlation score of each gene with the PCOS phenotype. This provides context for the gene ranking used in the GSEA, highlighting genes that are strongly associated with PCOS. The plot collectively illustrates the enrichment of the ribonucleoprotein complex biogenesis gene set within the ranked list of genes associated with PCOS, suggesting a potential biological relevance of this gene set to the development or progression of PCOS

Although this study is based on a dataset from a single center, the detailed clinical profiles reported in the original study provide important context which indicates that the lack of significant differences in age, BMI, FSH, or ART outcomes such as fertilization and cleavage rates suggests that both control and PCOS groups are largely comparable [8], thereby reducing potential confounding in transcriptomic comparisons reported in this present study. However, elevated AMH, LH, and TG levels in the PCOS group could reinforce the metabolic and endocrine differences identified in this study. This strengthens the biological relevance of findings reported in this study while also highlighting the importance of validating gene expression changes in an ethnically larger and clinically diverse population.

Conclusion

This study provides a transcriptomic insight into the molecular mechanisms underlying PCOS by identifying significant differentially expressed genes (DEGs) in cumulus cells, including CDH5, CLEC4D, and GNAT1. These genes are involved in critical pathways related to follicular development, insulin signaling, immune regulation, and oocyte competence. Despite originating from a single-center dataset, the original study reported no significant differences between PCOS and control subjects in age, BMI, or ART outcomes, supporting the validity of the transcriptomic comparisons. However, elevated

Table 2 Functional enrichment analysis of differentially expressed genes in PCOS

Source	Term name	Term ID	Intersections
GO:MF	Acyl binding	GO:0000035	GNAT1
GO:MF	Fibrinogen binding	GO:0070051	CDH5
GO:MF	Vascular endothelial growth factor receptor 2 binding	GO:0043184	CDH5
GO:MF	BMP receptor binding	GO:0070700	CDH5
GO:MF	Vascular endothelial growth factor receptor binding	GO:0005172	CDH5
GO:BP	Positive regulation of myeloid dendritic cell activation	GO:0030887	CLEC4D
GO:BP	Neural tissue regeneration	GO:0097719	GNAT1
GO:BP	Negative regulation of cyclic-nucleotide phosphodiesterase activity	GO:0051344	GNAT1
GO:BP	Regulation of myeloid dendritic cell activation	GO:0030885	CLEC4D
GO:BP	Protein localization to bicellular tight junction	GO:1,902,396	CDH5
GO:CC	Extrinsic component of plasma membrane	GO:0019897	CDH5, GNAT1
GO:CC	Side of membrane	GO:0098552	CDH5, CLEC4D, GNAT1
GO:CC	Dendritic spine head	GO:0044327	RGS7BP
GO:CC	External side of plasma membrane	GO:0009897	CDH5, CLEC4D
KEGG	Phototransduction	KEGG:04744	GNAT1
REAC	Activation of the phototransduction cascade	REAC:R-HSA-2485179	GNAT1
WP	Purinergic signaling	WP:WP4900	GNAT1
TF	Factor: KLF8; motif: NGGGGTGYGG	TF:M08818	CDH5, GNAT1, GPR25, RGS7BP
TF	Factor: NR4A2; motif: AGGTCANNNNNTGACCT	TF:M04478	GNAT1
CORUM	PAR-6-VE-cadherin complex, endothelial	CORUM:829	CDH5
CORUM	VEcad-VEGFR complex	CORUM:6586	CDH5
CORUM	ZO1-(beta)cadherin-(VE)cadherin-VEGFR2 complex	CORUM:5772	CDH5
HP	Electronegative electroretinogram	HP:0007984	GNAT1
HP	Congenital stationary night blindness with abnormal fundus	HP:0030639	GNAT1
GO:MF	Peptide hormone binding	GO:0017046	INSR, FSHR
GO:MF	Hormone binding	GO:0042562	INSR, FSHR
GO:MF	Follicle-stimulating hormone receptor activity	GO:0004963	FSHR
GO:BP	Female gonad development	GO:0008585	ESR1, INSR, FSHR
GO:BP	Development of primary female sexual characteristics	GO:0046545	ESR1,INSR,FSHR
GO:BP	Female sex differentiation	GO:0046660	ESR1, INSR, FSHR
GO:BP	Developmental growth	GO:0048589	ESR1, LEPR, INSR, FSHR
GO:BP	Development of primary sexual characteristics	GO:0045137	ESR1, INSR, FSHR
GO:BP	Uterus development	GO:0060065	ESR1, FSHR
GO:BP	Sex differentiation	GO:0007548	ESR1, INSR, FSHR
GO:BP	Reproductive structure development	GO:0048608	ESR1, INSR, FSHR
GO:BP	Reproductive system development	GO:0061458	ESR1, INSR, FSHR
GO:BP	Response to endogenous stimulus	GO:0009719	ESR1, LEPR, INSR, FSHR

Table 2 (continued)

Source	Term name	Term ID	Intersections
GO:BP	Reproductive process	GO:0022414	ESR1, LEPR, INSR, FSHR
GO:BP	Ovulation cycle process	GO:0022602	ESR1, FSHR
GO:BP	estrogen receptor signaling pathway	GO:0030520	ESR1, FSHR
GO:BP	Ovarian follicle development	GO:0001541	ESR1, FSHR
GO:BP	Ovulation cycle	GO:0042698	ESR1, FSHR
KEGG	Ovarian steroidogenesis	KEGG:04913	INSR, FSHR
HP	Delayed puberty	HP:0000823	ESR1, LEPR, INSR, FSHR
HP	Abnormality of the ovary	HP:0000137	ESR1, LEPR, INSR, FSHR
HP	Abnormal serum estradiol	HP:0025133	ESR1, LEPR, FSHR
HP	Abnormal circulating estrogen level	HP:0025132	ESR1, LEPR, FSHR
HP	Increased serum testosterone level	HP:0030088	INSR, FSHR
HP	Ovarian cyst	HP:0000138	ESR1, INSR, FSHR
HP	Enlarged polycystic ovaries	HP:0008675	ESR1, FSHR

This table presents the functional enrichment analysis results of the 10 sets of differentially expressed genes (*DEGs*) identified in polycystic ovary syndrome (*PCOS*). The table lists the enriched terms, their corresponding identifiers, and the genes that contribute to the enrichment. The terms are categorized by their ontology or database, including Gene Ontology Molecular Function (*GO:MF*), Gene Ontology Biological Process (*GO:BP*), Gene Ontology Cellular Component (*GO:CC*), Kyoto Encyclopedia of Genes and Genomes (*KEGG*), reactome (*REAC*), WikiPathways (*WP*), transcription factor (*TF*), comprehensive resource of mammalian protein complexes (*CORUM*), and Human Phenotype Ontology (*HP*). The "intersections" column indicates the specific genes from the input list that are associated with each term. This analysis aims to provide biological context and potential functional implications of the observed gene expression changes in PCOS

AMH, LH, and triglyceride levels in the PCOS group suggested underlying endocrine and metabolic disruptions as contributors to altered gene expression observed in this study. The findings suggest potential utility of cumulus cell transcriptomics in identifying biomarkers for PCOS-related infertility. However, the absence of functional validation and the limited ethnic diversity of the cohort are limitations necessitating future studies including qRT-PCR validation of top DEGs in independent populations, functional assays to determine gene roles in folliculogenesis, and multi-omics integration across diverse clinical phenotypes. These steps will be critical for translating molecular discoveries into diagnostic or therapeutic strategies for improving reproductive outcomes in women with PCOS.

Abbreviation

PCOS Polycystic ovary syndrome

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s43043-025-00241-w.

Additional file 1: Python Code for Gene Expression Analysis.

Acknowledgements
Not applicable

Authors' contributions
ABS: Conceptualization, Data Processing, Computational and Statistical Analysis, Writing of Original Draft. NMI: Formal Analysis, Validation, Writing of Original Draft, Writing, Review and Editing. KSA: Formal Analysis, Validation, Methodology. MAA: Writing, Review, Editing, and Validation. SSAE: Supervision, Validation, Writing, Review and Editing. KA: Writing, Review, and Editing

Funding
There is no funding support for this work.

Data availability
No datasets were generated or analysed during the current study.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Molecular, Cellular and Integrative Physiology Laboratory (MCIP), Department of Animal Science, Federal University of Agriculture Zuru, Zuru 872101, Kebbi State, Nigeria. [2]Laboratory of Vaccines and Biomolecules (VacBio), Institute of Biosciences, Universiti Putra Malaysia, Serdang, Seri Kembangan, Selangor 43400, Malaysia. [3]Department of Physiology, Biochemistry, and Pharmacology, College of Veterinary Medicine, University of Kirkuk, Kirkuk 36001, Iraq. [4]Department of Obstetrics and Gynaecology, Faculty of Health Sciences, College of Clinical Sciences, Ladoke Akintola University of Technology, Ogbomoso 210101, Nigeria. [5]Department of Animal Production, School of Food and Agriculture Technology, Federal University of Technology, Minna, Nigeria. [6]Department of Public Health, Fountain University, Osogbo, Nigeria.

Received: 20 April 2025 Accepted: 26 July 2025
Published online: 09 August 2025

References

1. Jiang B (2025) The global burden of polycystic ovary syndrome in women of reproductive age: findings from the GBD 2019 study. Int J Womens Health. https://doi.org/10.2147/IJWH.S490836
2. Sikiru AB, Adeniran MA, Akinola K, Behera H, Kalaignazhal G, Egena SSA (2023) Unraveling the complexity of the molecular pathways associated with polycystic ovary syndrome (PCOS) and identifying molecular targets for therapeutic development: a review of literature. Middle East Fertil Soc J 28:16
3. Azziz R, Carmina E, Chen Z, Dunaif A, Laven JSE, Legro RS, Lizneva D, Natterson-Horowtiz B, Teede HJ, Yildiz BO (2016) Polycystic ovary syndrome. Nat Rev Dis Primers 2:1–18
4. Stener-Victorin E, Teede H, Norman RJ, Legro R, Goodarzi MO, Dokras A, Laven J, Hoeger K, Piltonen TT (2024) Polycystic ovary syndrome. Nat Rev Dis Primers 10:27
5. Esencan E, Beroukhim G, Seifer DB (2022) Age-related changes in folliculogenesis and potential modifiers to improve fertility outcomes-a narrative review. Reprod Biol Endocrinol 20:156
6. Martinez CA, Rizos D, Rodriguez-Martinez H, Funahashi H (2023) Oocyte-cumulus cells crosstalk: new comparative insights. Theriogenology 205:87–93
7. Da Broi MG, Giorgi VSI, Wang F, Keefe DL, Albertini D, Navarro PA (2018) Influence of follicular fluid and cumulus cells on oocyte quality: clinical implications. J Assist Reprod Genet 35:735–751
8. Chen Y, Xie M, Wu S, Deng Z, Tang Y, Guan Y, Ye Y, He Q, Li L (2024) Multi-omics approach to reveal follicular metabolic changes and their effects on oocyte competence in PCOS patients. Front Endocrinol (Lausanne) 15:1426517. https://doi.org/10.3389/fendo.2024.1426517
9. Wickham H, Bryan J, Kalicinski M, Valery K, Leitienne C, Colbert B, Hoerl D, Miller E, Bryan MJ (2019) Package 'readxl.' Version 13:1
10. Yarberry W,(2021) Dplyr, CRAN Recipes: DPLYR, Stringr, Lubridate, and Regex in R. https://doi.org/10.1007/978-1-4842-6876-6_1 ISBN (Print) 978-1-4842-6875-9 ISBN (eBook) 978-1-4842-6876-6 Chapter Pages 1–58
11. Finotello F, Di Camillo B (2015) Measuring differential gene expression with RNA-seq: challenges and strategies for data analysis, Brief Funct. Genomics 14:130–142
12. Luo D, Wan X, Liu J, Tong T (2018) Optimally estimating the sample mean from the sample size, median, mid-range, and/or mid-quartile range. Stat Methods Med Res 27:1785–1805
13. Wickham H (2011) ggplot2, Wiley Interdiscip Rev. Comput Stat 3:180–185
14. Ranjan MK, Barot K, Khairnar V, Rawal V, Pimpalgaonkar A, Saxena S, Sattar AM (2023) Python: empowering data science applications and research. Journal of Operating Systems Development & Trends 10:27–33
15. Gu Z, Eils R, Schlesner M (2016) Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics 32:2847–2849
16. Innis SE, Reinaltt K, Civelek M, Anderson WD (2021) GSEAplot: a package for customizing gene set enrichment analysis in R. J Comput Biol 28:629–631
17. Kolberg L, Raudvere U, Kuzmin I, Adler P, Vilo J, Peterson H (2023) G:Profiler—interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). Nucleic Acids Res 51:W207–W212. https://doi.org/10.1093/nar/gkad347
18. Anazawa M, Ashibe S, Nagao Y (2025) Gene expression levels in cumulus cells are correlated with developmental competence of bovine oocytes. Theriogenology 231:11–20
19. Owens LA, Kristensen SG, Lerner A, Christopoulos G, Lavery S, Hanyaloglu AC, Hardy K, Andersen CY, Franks S (2019) Gene expression in granulosa cells from small antral follicles from women with or without polycystic ovaries. J Clin Endocrinol Metab 104:6182–6192
20. Khatun M, Lundin K, Naillat F, Loog L, Saarela U, Tuuri T, Salumets A, Piltonen TT, Tapanainen JS (2024) Induced pluripotent stem cells as a possible approach for exploring the pathophysiology of polycystic ovary syndrome (PCOS). Stem Cell Rev Rep 20:67–87
21. Ranjbaran A, Nejabati HR, Ghasemnejad T, Latifi Z, Hamdi K, Hajipour H, Raffel N, Bahrami-Asl Z, Hakimi P, Mihanfar A (2019) Follicular fluid

levels of adrenomedullin 2, vascular endothelial growth factor and its soluble receptors are associated with ovarian response during ART cycles. Geburtshilfe Frauenheilkd 79:86–93
22. Xie Q, Hong W, Li Y, Ling S, Zhou Z, Dai Y, Wu W, Weng R, Zhong Z, Tan J (2023) Chitosan oligosaccharide improves ovarian granulosa cells inflammation and oxidative stress in patients with polycystic ovary syndrome. Front Immunol 14:1086232
23. Liu H, Tang J, Du Y, Saadane A, Samuels I, Veenstra A, Kiser JZ, Palczewski K, Kern TS (2019) Transducin1, phototransduction and the development of early diabetic retinopathy. Invest Ophthalmol Vis Sci 60:1538–1546
24. Sayutti N, Abu MA, Ahmad MF (2022) PCOS and role of cumulus gene expression in assessing oocytes quality. Front Endocrinol (Lausanne) 13:843867
25. Huang Z, Wells D (2010) The human oocyte and cumulus cells relationship: new insights from the cumulus cell transcriptome. Mol Hum Reprod 16:715–725
26. Mihailovic MK, Chen A, Gonzalez-Rivera JC, Contreras LM (2017) Defective ribonucleoproteins, mistakes in RNA processing, and diseases. Biochemistry 56:1367–1382
27. Chakravarthi VP, Ratri A, Masumi S, Borosha S, Ghosh S, Christenson LK, Roby KF, Wolfe MW, Rumi MAK (2021) Granulosa cell genes that regulate ovarian follicle development beyond the antral stage: the role of estrogen receptor β. Mol Cell Endocrinol 528:111212
28. Roy N, Mascolo E, Lazzaretti C, Paradiso E, D'Alessandro S, Zaręba K, Simoni M, Casarini L (2021) Endocrine disruption of the follicle-stimulating hormone receptor signaling during the human antral follicle growth. Front Endocrinol (Lausanne) 12:791763

Publisher's Note