ORIGINAL ARTICLE

Integrated bioinformatics analysis and screening of hub genes in polycystic ovary syndrome

Gan Qiao[1] · Jinshan Xing[2] · Xin Luo[1] · Chunxiang Zhang[3] · Jingyan Yi [4]

Abstract

Purpose Polycystic ovary syndrome (PCOS) is one of the most common endocrine and metabolic disorders, posing a serious threat to the health of women. Herein, we aimed to explore new biomarkers and potential therapeutic targets for PCOS by employing integrated bioinformatics tools.

Methods Three gene expression profile datasets (GSE138518, GSE155489, GSE106724) were obtained from the Gene Expression Omnibus database and the differentially expressed genes in PCOS and normal groups with an adjusted p-value < 0.05 and a |log fold change (FC)| > 1.2 were first identified using the DESeq package. The weighted correlation network analysis (WGCNA) R package was used to identify clusters of highly correlated genes or modules associated with PCOS. Protein-protein interaction (PPI) network analysis and visualization of genes in the key module were performed using the STRINGdb database and the NetworkX package (edge > 5), respectively. The genes overlapping among the key module genes and PCOS-associated genes were further analyzed. Ligand molecules with strong binding energy < -10 kJ/mol to GNB3 were screened in the drug library using MTiOpenScreen. AutoDock, ChimeraX, and BIOVIA Discovery Studio Visualizer were further used to elucidate the mechanism of ligand interaction with GNB3. Finally, the relationship between GNB3 and PCOS was verified using experimental models in vivo and in vitro.

Results Of the 11 modules identified by WGCNA, the black module had the highest correlation with PCOS (correlation $= 0.96$, $P = 0.00016$). The PPI network of 351 related genes revealed that *VCL, GNB3, MYH11, LMNA, MLLT4, EZH2, PAK3*, and *CHRM1* have important roles in PCOS. The hub gene *GNB3* was identified by taking the intersection of PCOS-related gene sets. MTiOpenScreen revealed that five compounds interacted with GNB3. Of these five, compound 1 had the strongest binding ability and can bind amino acids in the WD40 motif of GNB3, which in turn affects the function of the G protein-coupled receptor β subunit. GNB3 was also significantly downregulated in PCOS models.

Conclusion We identified the hub gene *GNB3* as the most important regulatory gene in PCOS. We suggest that compound 1 can target the WD40 motif of GNB3 to affect related functions and must be considered as a lead compound for drug development. This study will provide new insights into the development of PCOS-related drugs.

Keywords PCOS · WGCNA · Molecular docking · GNB3 · Bioinformatics

Abbreviations

PCOS	Polycystic ovary syndrome
GCs	Granulosa cells
DEGs	Differentially expressed genes
BMI	Body mass index
WGCNA	Weighted gene co-expression network analysis
GEO	Gene Expression Omnibus
GO	Gene ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
PPI	Protein–Protein Interaction.

These authors contributed equally: Gan Qiao, Jinshan Xing, and Xin Luo

✉ Chunxiang Zhang
    zhangchx999@163.com

✉ Jingyan Yi
    jingyany@swmu.edu.cn

Extended author information available on the last page of the article

Introduction

Polycystic ovary syndrome (PCOS) is one of the most common endocrine and metabolic disorders, characterized by polycystic ovaries, hyperandrogenemia, hypermenorrhea, hyperinsulinemia, chronic anovulation, and metabolic

disorders [1–3]. PCOS has a complex pathophysiology that affects approximately 6–10% of women of reproductive age, and also leads to an increased risk of diabetes and cardiovascular disease, posing a serious health risk to women [4, 5].

PCOS often presents in familial clusters, suggesting that genetic factors play an important role in PCOS [6]. Although great efforts have been made in genome-wide association studies for decades, only a few genetic variants have been identified in different populations, which account for nearly 10% of the heritability of PCOS, and the molecular genetic mechanism of PCOS is still unclear [7, 8]. As a component of the ovary, granulosa cells (GCs) surround the oocyte and have a role in maintaining the intracellular environment. Dysfunction of GCs may lead to abnormal follicle development [9].

Unlike previous methods of analyzing differentially expressed genes (DEGs) between samples, weighted gene co-expression network analysis (WGCNA) focuses on correlations between genes, which can reduce bias and subject judgment [10, 11]. The advantage of WGCNA is that it can quickly filter out the set of genes associated with a specific phenotype from massive data and identify hub genes that play an important role in transcriptional regulation [12].

This study aimed to screen novel biomarkers in GCs associated with PCOS using WGCNA-based screening. We first screened for PCOS-associated core genes in the RNA-seq datasets. Furthermore, we used molecular docking to identify small molecule compounds that may act on the hub gene. Finally, we checked the levels of the hub gene expression in the PCOS models in vitro and in vivo. Our study highlights the importance of key genes in the etiology of PCOS and provides a basis for subsequent small molecule drug development.

Materials and methods

Data collection

GCs provide nutrients and growth regulators for oocyte development and play an important role in PCOS development. We obtained the GSE138518, GSE155489, and GSE106724 gene expression datasets from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) for studies of GCs in PCOS with 10 controls (control group) and 14 PCOS patients (case group). The NCBI database (https://www.ncbi.nlm.nih.gov/gene) and OMIM database (https://omim.org) were used to screen PCOS disease genes [13, 14]. Ethical approval was obtained to report on the patients involved, as all databases used in this study are public databases.

The three datasets contained PCCS patients with similar features. The patients included in the GSE138518 dataset were diagnosed with PCOS by clinicians by at least two of the three Rotterdam criteria. The control group included fertile or infertile women with tubal blockages and with normal menstrual cycles, no polycystic ovary morphology, or other chronic diseases. All participants were non-related Han people aged between 20 and 35 years [15]. The patients included in the GSE155489 dataset were diagnosed with PCOS based on the Rotterdam criteria. The control group had regular menstrual cycles, normal sonographic appearance of ovaries, and no diabetes or clinical signs of PCOS. The study subjects included the control group with body mass index (BMI) between 17.70 and 23.50 and PCOS patients with BMI between 19.00 and 28.10 [16]. The average age of the PCOS patients and controls was 32.40 ± 1.29 and 35.60 ± 2.23, respectively. The patients included in the GSE106724 dataset were diagnosed by clinicians based on the Rotterdam criteria. All controls were women with regular menstrual cycles and sex hormone levels within normal ranges. All patients were aged younger than 35 years. For each patient, ovarian morphology was assessed by ultrasound to exclude structural abnormalities [17].

Screening of DEGs

Data from the GEO database were used for determining gene expression levels based on the number of reads [18]. The read counts were normalized using the DESeq package for differential expression analysis. Genes were considered statistically significant at p-value < 0.05 and a |log fold change (FC)| > 1.2. DEGs were visualized using the pheatmap package and EnhancedVolcano package in R.

Functional enrichment analysis

Gene ontology (GO) analysis was used to identify the possible molecular function of and visualize the potential biological significance of the large list of genes [19]. Kyoto Encyclopedia of Genes and Genomes (KEGG) database (http://www.genome.ad.jp/kegg/) was used to analyze the potential functions of the identified DEGs [20]. Enrichment analysis was conducted using the DisGeNET (https://www.disgenet.org/) [21] database and clusterProfiler package. KEGG enrichment analysis was conducted using the TCGAbiolinker API package. The biological process terms with a p-value < 0.05 were considered statistically significant.

WGCNA analysis for the identification of hub genes

Analysis of differentially expressed mRNAs was performed using the R software based on the criteria of |Fold Change|

>1.2 and error detection rate (FDR) < 0.05 [22]. WGCNA is a systems biology method for describing correlation patterns between genes across multiple samples [11, 12]. Differentially expressed mRNAs were screened for determining the modules most relevant to PCOS using the WGCNA package. The correlation matrix (Sij) was converted to an adjacency matrix (Aij) based on soft threshold β that can approximate the scale-free distribution ($R^2 > 0.8$). This transformation allowed us to build networks with higher biological signals, which is the focus of the WGCNA approach. A topological overlap matrix was used for the visualization of a simplified network diagram for module identification. The hierarchical clustering tree formed by average linkage hierarchical clustering was also used in the determination of modules. To identify the key modules closely related to PCOS, module eigengene, gene significance, module membership, and other parameters were calculated using different settings in the WGCNA model. The genes in the key modules with high correlation and signification were further analyzed using disease enrichment and KEGG enrichment analysis to explore the biological functions of the genes in the key modules. Genes from the enriched pathway were crossed over with genes in the PCOS-related modules to determine hub genes for further analysis.

Construction of protein-protein interaction (PPI) network

A PPI network was generated using the STRING database (https://string-db.org), limiting the category to "*Homo sapiens*" [23]. The degree of node > 3 was selected for the PPI network., and the networkx package in python 3 was used to visualize the PPI network.

Molecular docking

The number of edges in a PPI network represents the interactions of a protein, therefore, by counting the edges, we identified GNB3 as the protein most associated with PCOS. A library of compounds was screened for identifying the target compounds of GNB3. MTiOpenScreen (http://bioserv.rpbs.univ-paris-diderot.fr/services/MTiOpenScreen/) enables researchers to apply virtual screening using different chemical libraries on traditional or more challenging protein targets such as protein-protein interactions [24]. Autodock is a widely used molecular docking software that allows conformational changes in small molecules to bind free energy as a basis for evaluating docking results [25]. BIOVIA Discovery Studio Visualizer is a feature-rich molecular modeling application for viewing, sharing, and analyzing protein and small molecule data [26]. ChimeraX (https://www.cgl.ucsf.edu/chimerax/) is a molecular visualization tool that allows hydrogen bonding analysis, atomic contacts, and sequence presentation [27]. AutoDock, ChimeraX, and BIOVIA Discovery Studio Visualizer were used to elucidate the mechanism of ligand interaction with GNB3. The binding strength between GNB3 and compounds was evaluated based on binding energy and hydrogen bonding analysis [28].

Cell Culture and development of PCOS cell model

KGN cells, a steroidogenic human granulosa cell-like tumor cell line was purchased from iCell Bioscience Inc (China) and identified by short tandem repeat profiling. Cells were maintained in Dulbecco's modified Eagle medium F-12 supplemented with 10% fetal bovine serum and penicillin/streptomycin (100 units/ml) at 37 °C and 5% CO_2. Cells were stimulated with DHEA (Macklin, China) for 48 h to simulate the PCOS model in vitro.

Construction of PCOS mouse model

A PCOS mouse model was established in 23-day-old C57BL/6 female mice by daily subcutaneous injection of 60 mg/kg DHEA (Macklin, China) for 21 days [29]. All animal procedures were conducted under the approval of the Animal Care and Use Committee of Southwest Medical University. On the last day of injection, the mice were executed by cervical dislocation and ovarian tissues were collected and set aside.

Western blot analysis

Fresh cells were lysed with RIPA lysis buffer. Protein was separated by SDS-PAGE and transferred onto PVDF membranes. The membranes were incubated in primary antibodies against GNB3 (PU201127,1:3000) and β-actin (66009-1-Ig,1:10000). HRP-conjugated antibodies against mouse or rabbit (1:10000, Proteintech Group, Inc.) were used as the secondary antibodies. Immunoblot imaging was performed using the BIO-RAD ChemiDoc™ XRS + Molecular Imager. The western blot was normalized to β-actin.

RNA extraction and RT-qPCR analysis

Total RNA was isolated using TRIzol reagent (Invitrogen). RNA was synthesized into cDNA using the HiScript II 1st Strand cDNA Synthesis Kit (Vazyme). Taq Pro Universal SYBR qPCR Master Mix (Vazyme) was used to perform quantitative PCR (qPCR). QuantStudio™ Design & Analysis Software was used to analyze the samples. Gene expression was normalized to that of *ACTB*. The following primers were used:
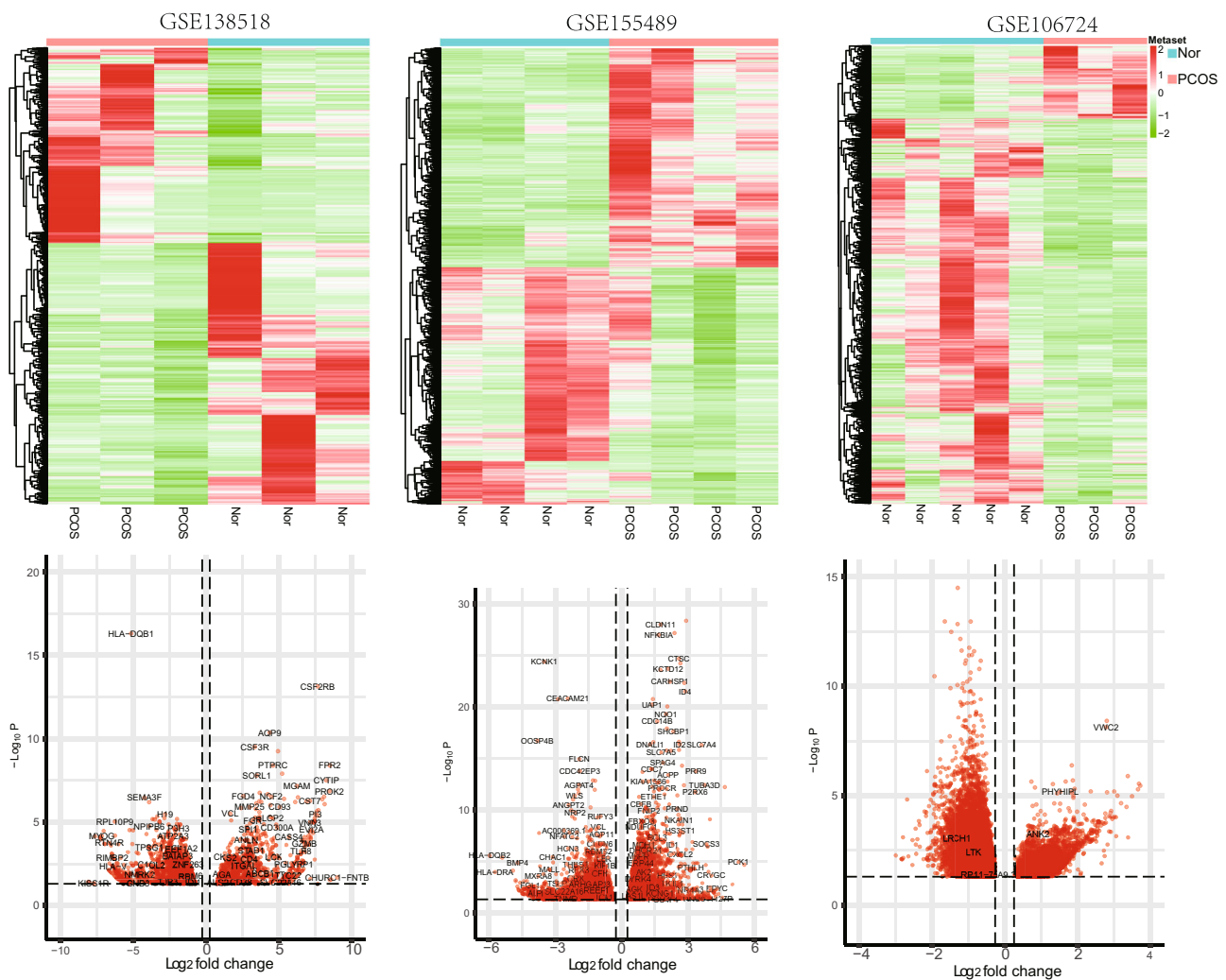
Fig. 1 The GEO database for differentially expressed genes of GCs in PCOS. Differential expression genes of GSE138518, GSE155489, GSE106724 was shown in the heatmap (up panel) and Volcano plot (down panel). The red cell was the upregulation gene and the green was the downregulation gene in the heatmap, the expression signification was shown as log10 p values in the y axis and the fold change was shown in the x-axis in the Volcano plot

β-actin(H):
5′-CATGTACGTTGCTATCCAGGC-3′ (Forward)
5′-CTCCTTAATGTCACGCACGAT-3′ (Reverse)
β-actin(M):
5′-GGCTGTATTCCCCTCCATCG-3′ (Forward)
5′-CCAGTTGGTAACAATGCCATGT-3′ (Reverse)
GNB3(H):
5′-CGGACGTTAAGGGGACACC-3′ (Forward)
5′-CGAGGCACTTACCAGCAGC-3′ (Reverse)
GNB3(M):
5′-AAGAAGCAGATTGCTGATGCC-3′ (Forward)
5′-GTCCCCTTAATGTCCTCCGTG-3′ (Reverse)

Statistical analyses

All data are expressed as mean ± SEM. Statistical analysis was performed using the GraphPad Prism 8 software.

Differences between two independent groups were calculated using Unpaired Student's t-test and one-way ANOVA with Tukey's multiple-comparisons test. P-values < 0.05 were considered statistically significant and are denoted as follows: $* < 0.05$, $** < 0.01$, and $*** < 0.001$.

Results

Annotation of DEGs in GCs of PCOS patients and KEGG pathway analysis

As shown in Fig. 1, the GSE138518, GSE155489, and GSE106724 datasets contained 2043 (1419 mRNA and 624 ncRNA), 5257 (3843 mRNA and 1414ncRNA) and 16,537 (6369 mRNA and 10,168 ncRNA) DEGs. Of the 5257 DEGs in GSE155489, 2751 were up-regulated and 2506 were
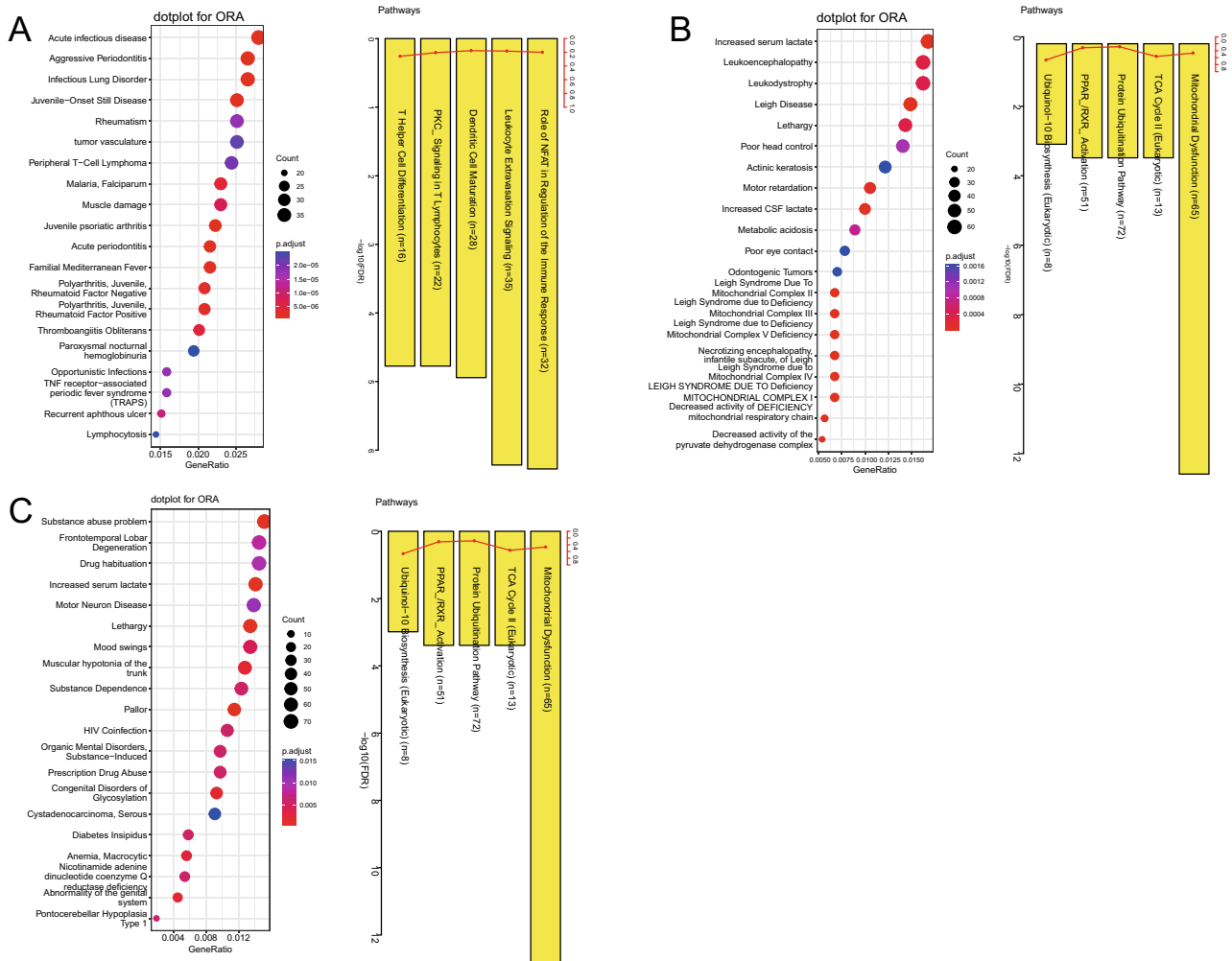
Fig. 2 DisGeNET and KEGG analyses of regulated genes in PCOS vs. Normal. **A** DisGeNET and KEGG enrichment analysis of DGEs in PCOS vs. Normal in GSE138518, **B** GSE155489, **C** GSE106724

down-regulated (Fig. 1). Of the 16,537 DEGs in GSE106724, 2673 were up-regulated and 13,864 were down-regulated (Fig. 1). Next, we performed DisGeNET and KEGG analyses of the DEGs in each dataset and found that functionally, DEGs were mainly enriched in acute infectious diseases, increased serum lactate, and increased serum lactate (Fig. 2A–C). Pathway analysis revealed that DEGs were mainly enriched in the synthesis, secretion, and action of the growth hormone and the metabolism of fructose and mannose (Fig. 2A–C). These data corroborate the fact that PCOS is a hormone-regulated metabolic disease closely related to inflammation.

Construction of WGCNA modules using PCOS gene sets

Due to the scale-free indexes of the GSE106724 and GSE138518 datasets have a non-scale-free network distribution, we choose the GSE155489 dataset for WGCNA analysis. After excluding the missing and abnormal values, 4731 mRNAs were further analyzed. For WGCNA analysis, the soft threshold power was selected as 11 to ensure a scale-free network distribution (scale-free index $R2 > 0.9$; connectivity = 1) (Fig. 3A). At this point, the network conforms to the power-law distribution, which is closer to the real biological state. The gene dendrogram and corresponding module colors are shown in Fig. 3B. A total of 12 modules were identified by the Dynamic Tree Cut algorithm (Fig. 3B), and 596, 595, 566, 559, 555, 552, 524, 504, and 282 DEGs were enriched in the blue, black, yellow, magenta, red, purple, turquoise, pink, brown, green and green module, respectively. The phenotypic gene set heat map, revealing the correlation between module genes and PCOS, yielded the highest correlations for the black and pink modules (Fig. 3C). The correlation value and p-value of the regression curve in the module membership vs. gene significance curve ($r = 0.96$, $P < 0.01$) also suggested that genes in the black modules may be correlated with PCOS (Fig. 3D). We extracted the 421 relevant genes and DEGs
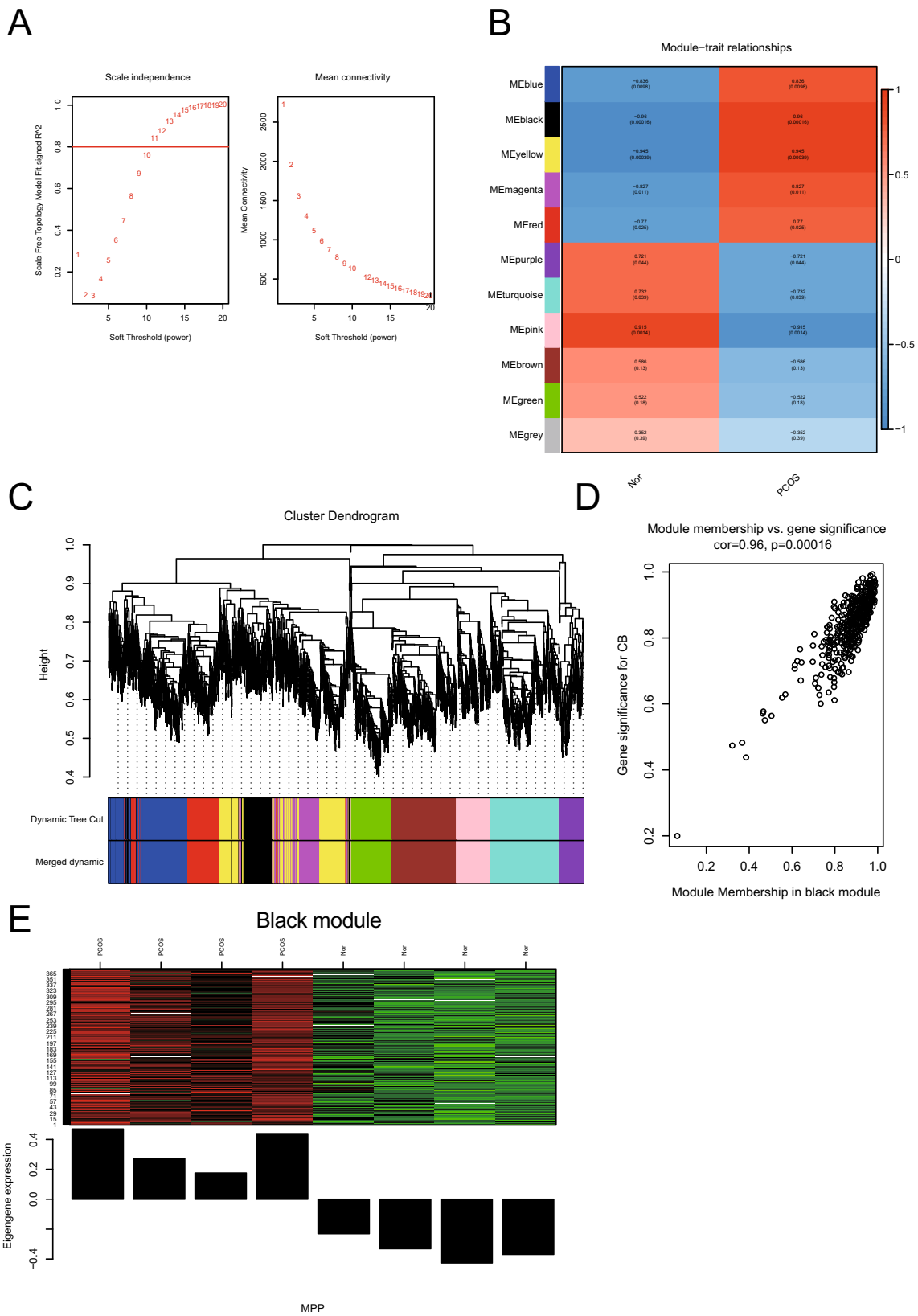
Fig. 3 Identification of PCOS-related modules by WGCNA. **A** The impact of soft-threshold power on the scale-free topology fit index curve was shown in the left panel; the right panel displays the impact of soft-threshold power on the mean connectivity. **B** Gene clustering tree (dendrogram) obtained by hierarchical clustering of adjacency-based dissimilarity. **C** The module-phenotype relationships. Each row corresponds to a module eigengene, column to the phenotype (Normal or PCOS). The correlation and *p*-value were also shown in the cells; red for positive correlation and blue for negative correlation. **D** The Module membership vs. gene significance curve indicated the black module gene is the most correlation with the PCOS. **E** Black module gene dataset shown that the 421 genes in the module, which were relation with the PCOS. Threold of FilterGenes was identified by abs (GS1) > 0.9 and abs (datKME$MM.black) > 0.9. Analysis was performed using R software based on the criteria of | Fold Change | > 1.2 and error detection rate (FDR) < 0.05
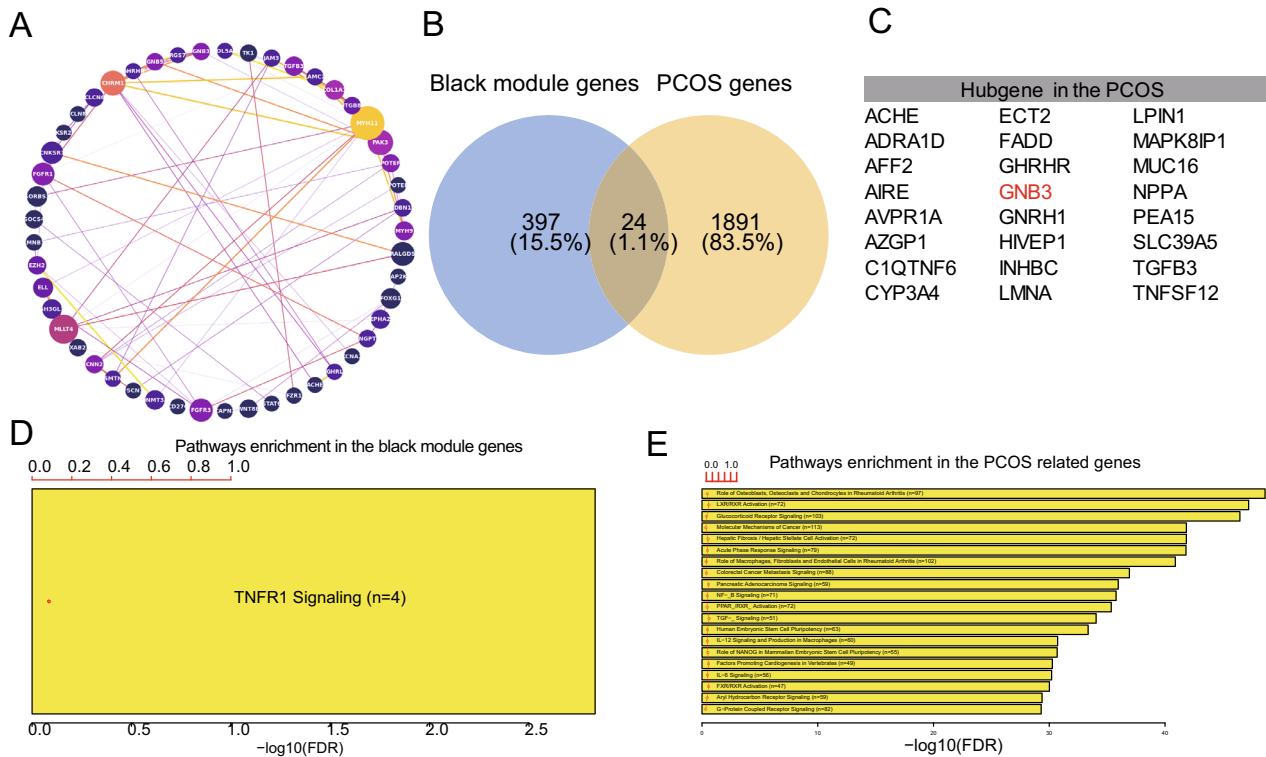

Fig. 4 Screening of core hub gene in PCOS. **A** The hub genes were identified from the black module gene dataset by PPI. **B** The Venn diagram of PCOS disease gene dataset and black module gene dataset. **C** Hub-genes in PCOS were shown. GNB3 marked in red is the core hub gene with PCOS. **D** The gene enrichment analysis of the black module gene dataset. **E** The gene enrichment analysis of the PCOS disease gene which is annotation in the OMMI database and Gene-cards database. The biology process terms with a p-value < 0.05 were considered statistically significant

expression profiles from the black module for further analysis. We found that the module gene set was significantly and positively correlated with the development of PCOS, and should be further investigated (Fig. 3E).

Screening for hub genes in PCOS

The STRINGdb database was used to identify key genes using the NetworkX package in python for PPI analysis and visualization of genes in the black module. There were 38 nodes in the PPI network representing protein interactions (Fig. 4A). Genes overlapping in the black module and those related to PCOS in previous reports are shown in Fig. 4B, and 24 related genes were identified (Fig. 4C). *GNB3* with a high degree (nodes) in the network was identified to be associated with PCOS. Previous reports have shown that

GNB3 overexpression is associated with obesity and metabolic syndrome, and insulin resistance and high adiposity and metabolic syndrome are known features of PCOS [30]. Considering the functional diversity (number of edges) in the PPI network, we hypothesize that *GNB3* may be an important regulatory gene in PCOS (Fig. 4A, C). We also performed enrichment analysis of the genes associated with the black module as well as those that play a role in PCOS. We found that genes associated with the black module were significantly associated with the TNFR1 signaling pathway (Fig. 4D), while PCOS-related genes as determined by genecards and the Catalog of Human Genes and Genetic Disorders database (OMIM) were mainly significantly associated with the inflammatory and metabolic disorder pathway (Fig. 4E). The above results suggest that *GNB3* may be an important inflammatory regulator in PCOS.
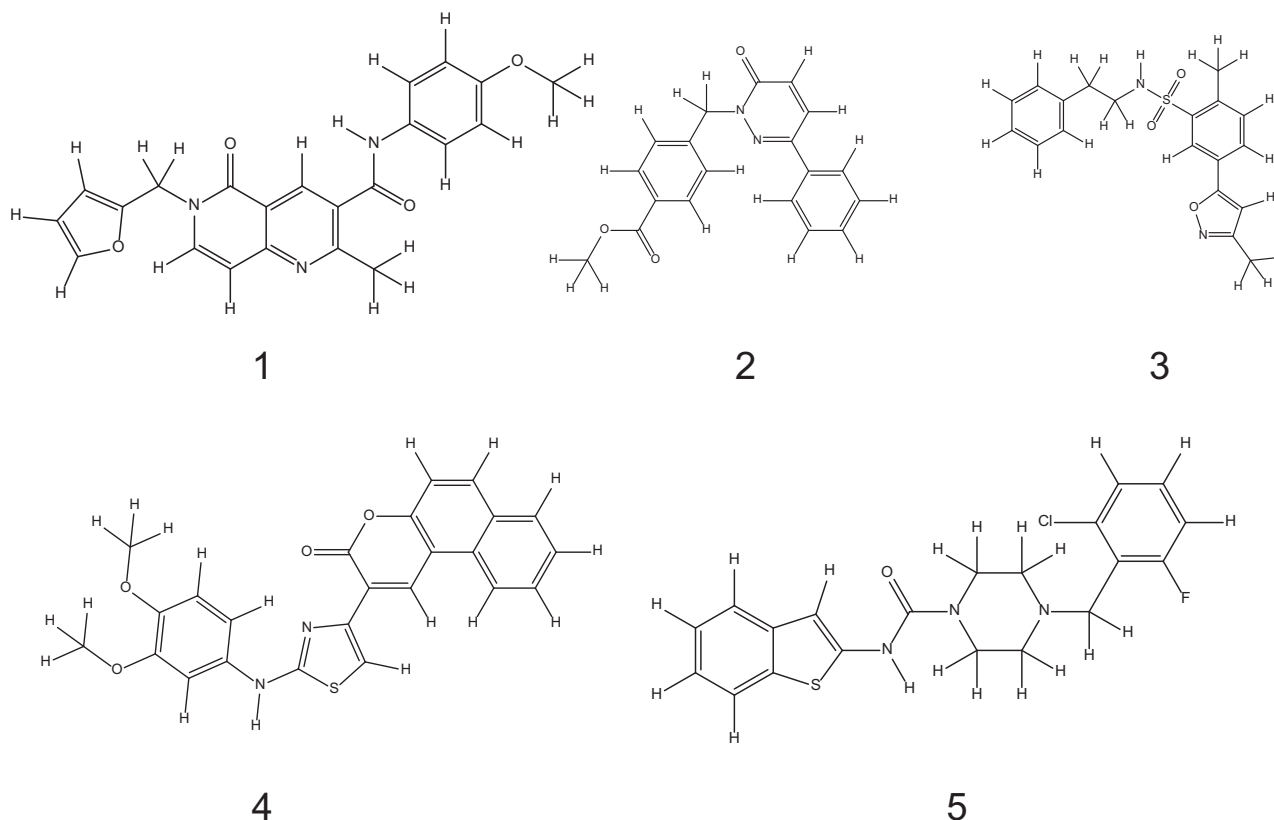
Fig. 5 Structures of five compounds bound to GNB3. Five compounds were: (1) 6-(furan-2-ylmethyl)-N-(4-methoxyphenyl)-2-methyl-5-oxo-1,6-naphthyridine-3-carboxamide, (2) methyl 4-[(6-oxo-3-phenylpyr-idazin-1(6H)-yl)methyl]benzoate, (3) 2-methyl-5-(3-methyl-1,2-oxazol-5-yl)-N-(2-phenylethyl) benzenesulfonamide, (4) 2-[2-(3,4-dimethoxyanilino)-1,3-thiazol-4-yl]benzo[f] chromen-3-one, (5) N-(1-benzothiophen-2-yl)-4-[(2-chloro-6-fluorophenyl)methyl]piperazine-1-carboxamide

Table 1 The feature and the binding energy of ligands that binding with GNB3

Compound (SMILES)	Model ID	Energy	LogP	MW
COc1ccc(NC(= O)c2cc3c(ccn(Cc4ccco4)c3 = O)nc2C)cc1	1	−10.4	2.28	389.404
COC(= O)c1ccc(Cn2nc(ccc2 = O)-c2ccccc2)cc1	2	−10.3	2.79	320.3419
Cc1cc(on1)-c1ccc(C)c(c1)S(= O)(= O)NCCc1ccccc1	3	−10.2	3.72	356.4387
COc1ccc(Nc2nc(cs2)-c2cc3c(ccc4ccccc34)oc2 = O)cc1OC	4	−10.1	5.68	430.4757
Fc1cccc(Cl)c1CN1CCN(CC1)C(= O)Nc1cc2ccccc2s1	5	−10	4.5	403.9008

Screening of small molecule compounds targeting GNB3

GNB3 forms the β3 subunit of G proteins and has a major regulatory role in the function of G protein-coupled receptors [29, 30]. We constructed the 3D structure of GNB3 using alphafold. Then, the possible binding pockets of GNB3 were evaluated by Ghecom (http://biosig.unimelb.edu.au/). We chose the top-ranked pocket locations and obtained the center of the pockets as −3.09, 3.976, and −2.765 points (x, y, z). We screened the molecules binding GNB3 using MTiOpenScreen and selected five compounds with binding energy <10 kJ/mol according to the change of ligand binding energy (Fig. 5 and Table 1).

Conformational analysis of GNB3 and its ligand

Analysis of the 3D structure of GNB3 suggested that it has a β-helical WD40 superprotein structure with seven blades (Fig. 6A). Autodock analysis in the PyRx software was used to determine the structure of the complex formed between compound 1 and GNB3 [31]. Visualization revealed that compound 1 was bound to the central region of the WD40 motif superstructure of GNB3 through hydrogen bonding, hydrophobic bonding, and electrophilic interaction (Fig. 6B). In particular, compound 1 interacts with LEU190, VAL276, and SER316 of GNB3 via hydrogen bonding and with ALA60, ARG150, LEU190, CYS233, and CYS317 of GNB3 via hydrophobic and
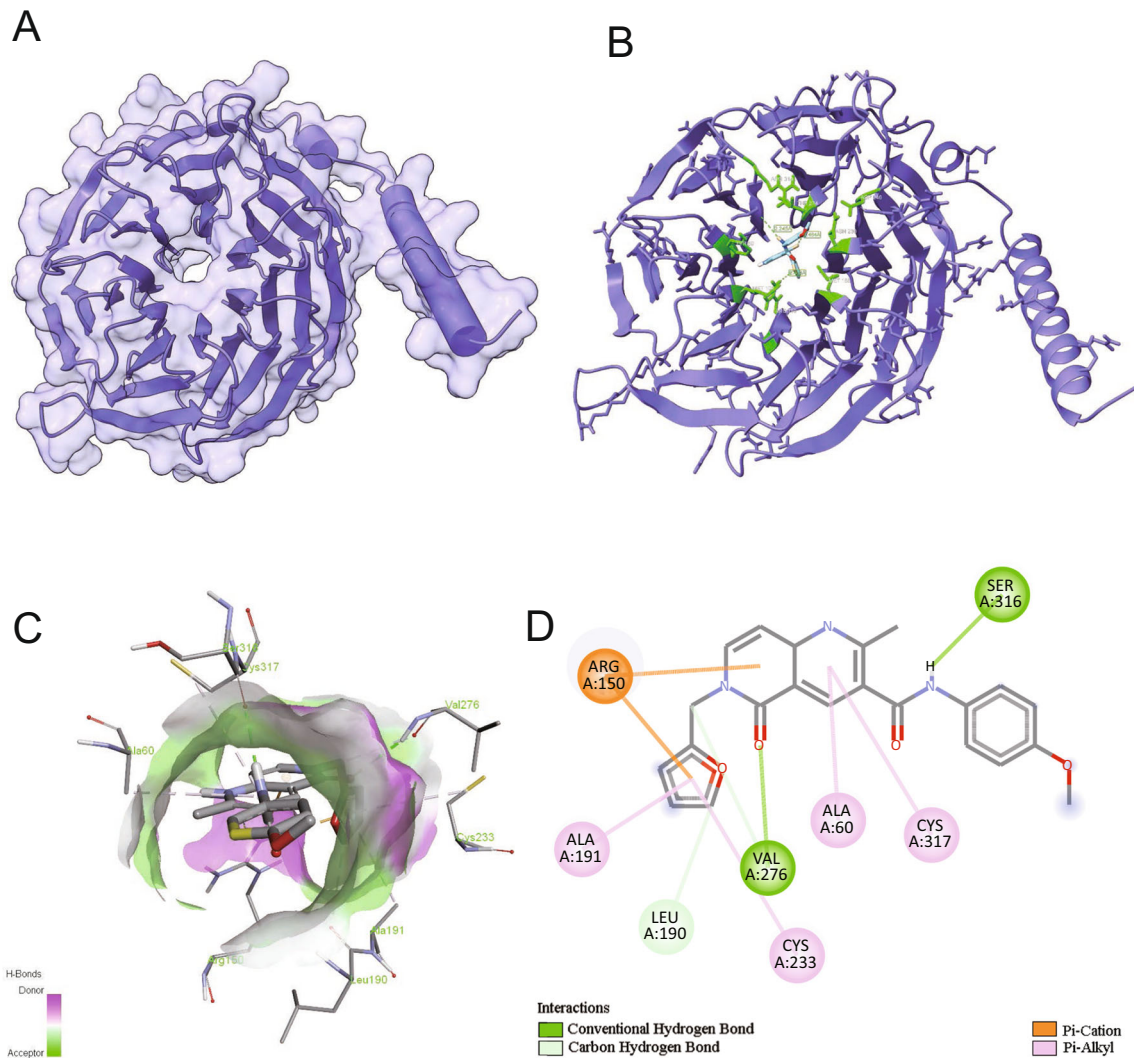
Fig. 6 Molecular docking of the GNB3 with the compound 1. **A** The conformation of GNB3 was shown with the second structure Helices, Sheets and strands. **B** The conformation of GNB3-ligand complex was shown with the H-bond (green dash lines) and atom distance (the number). GNB3 was shown in the post docking analysis visualized by Discovery Studio Visualizer in both 3D (**C**) and 2D (**D**) models

electrophilic interactions (Fig. 6C, D). These results suggest that compound 1 may serve as a lead compound for targeting GNB3 drugs and has important potential in the development of new drugs for PCOS.

GNB3 is significantly downregulated in PCOS

To verify whether *GNB3* is a hub gene in PCOS, we established PCOS models in vivo and in vitro. To reduce heterogeneity, we collected mouse ovaries for follow-up testing in a combined sample of the control and PCOS groups. The results showed that *GNB3* mRNA levels were significantly reduced in the PCOS mouse model (Fig. 7A). The expression of GNB3 in KGN cells was also decreased after DHEA treatment (Fig. 7B). Consistent with RNA levels, the protein levels of GNB3 were significantly

reduced in the PCOS model (Fig. 7C, D). These results suggest that GNB3 is significantly associated with and may be valuable in the study of PCOS.

Discussion

WD40 can assemble into different complexes by interacting with multiple proteins, DNA, or RNA. These complexes regulate a variety of important biological functions in living organisms, including DNA damage repair, DNA replication, histone methylation site recognition, phosphorylation site recognition, ubiquitinated substrate recognition, assembly of apoptotic vesicles, vesicle formation, cell signaling, and nucleolar pore formation [31–33]. Moreover, many mutated sites on the WD40 domain proteins have

Fig. 7 GNB3 is significantly downregulated in PCOS. qRT-PCR was performed to detect the expression of GNB3 in PCOS mouse models (**A**) and DHEA-treated KGN cells (**B**). Western blot analysis was performed to detect the expression of GNB3 in PCOS mouse models (**C**) and DHEA-treated KGN cells (**D**). β-actin was used as a normalization standard. Data are shown as mean ± S.E.M. for three independent experiments (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ by Student's t-test)
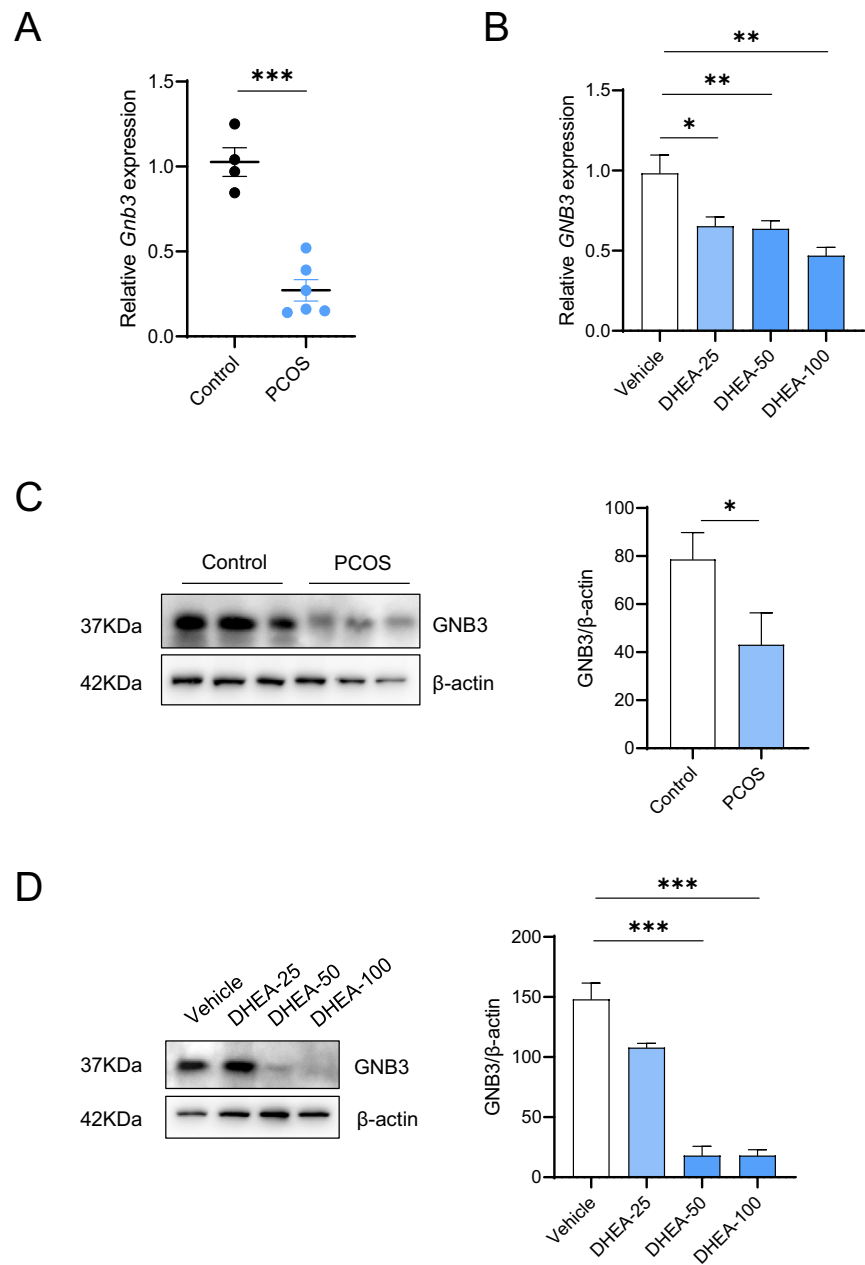
been shown to be associated with several human diseases [34, 35].

GNB3 encodes the G protein β3 subunit. In vivo, the specific interactions of the Gβ subunit with the Gα and Gγ subunits have not yet been clarified [30, 36]. Mutations in GNB3 (C825T) have been shown to be located in exon 10 of GNB3, and although they do not change the amino acid sequence, the T allele is associated with variable splicing in exon 9, which can affect various diseases such as hypertension and obesity [37, 38]. Therefore, GNB3 may be closely related to metabolic diseases. PCOS is one of the most common endocrine and metabolic disorders with a complex pathophysiology and poses a serious threat to the

health of women [39, 40]. The pathogenesis of PCOS remains largely unknown; follicles in patients with PCOS are arrested during the small sinus follicular phase and are not recruited to develop into mature follicles, thus stopping ovulation [9]. GCs in the follicles provide nutrients and growth regulators for oocyte development, which is essential for follicle development and ovulation [9]. Therefore, the analysis of GCs is important in PCOS research. GNB3 is associated with obesity and metabolic syndrome, and the essence of PCOS is metabolic abnormalities. Therefore, the selective targeting of drugs to GNB3 may be valuable in the treatment of PCOS. In our study, we found that compound 1 interacts with GNB3 mainly through hydrogen bonds and

hydrophobic bonds, which may affect its binding to Gα and Gγ. It has been shown that Aβ42 is an FPR2 receptor agonist, and analysis of the structural complexes of FPR2 and G proteins revealed that the Gβ subunit (GNB1) has direct interactions with Gα as well as Gγ [41]. GNB3 and GNB1 belong to the Gβ family with similar amino acid sequences and may have close structural functions. Therefore, we speculate that compound 1 may function as an agonist of GNB3, but this needs further validation.

WGCNA is used to analyze the association between genes and cluster them into modules that depict the complex relationships in the network of gene regulation [12]. NcRNAs play an important role in PCOS development. Therefore, analysis of the correlation between these modules and sample phenotypes can reveal the molecular characteristics of specific phenotypes, which is more advantageous than the DEGs method [42]. In this study, the WGCNA method was used to screen the gene modules significantly associated with PCOS and functional enrichment analysis of the genes in these modules was then conducted. The hub gene *GNB3* was mined to PCOS in association with the disease gene set, and molecular docking was used to screen for compounds binding GNB3. We also assessed the expression of GNB3 in PCOS models. We found that GNB3 was significantly downregulated in the PCOS models, suggesting that GNB3 may be a potential target in PCOS. Our study is limited in that we only determined the differential phenotypic expression of GNB3 in PCOS, and its functional role needs further validation.

Nevertheless, we used integrated bioinformatics tools to screen for GNB3, the core hub gene of PCOS, and uncover the most promising target compound, providing important clues in the development of new drugs for PCOS.

Conclusions

Based on public databases and multiple computer algorithms, we identified *GNB3* as a core hub gene in PCOS, and suggest that it may be an important regulator of inflammation in PCOS. Analysis using MTiOpenScreen revealed five compounds from the drug library that interacted with GNB3 and finally selected compound 1 with the strongest ability to bind GNB3. Compound 1 was found to bind amino acids in the WD40 motif of GNB3 forming the β-subunit, which in turn affects the function of the G protein-coupled receptor. Meanwhile, the in vivo and in vitro analysis confirmed that GNB3 was significantly downregulated in PCOS models. Therefore, we postulate that GNB3 is associated with PCOS and its ligands must be considered for the development of drugs for PCOS. This study will provide new insights into the development of PCOS-related drugs.

Data availability

The data that support the findings of this study are available online. Contact with correspondence author for any data and material on reasonable request.

Compliance with ethical standards

Conflict of interest The authors declare no competing interests.

Ethics approval and consent to participate Ethical approval was obtained to report on the patients involved as all databases used in this study are public databases. The animal experiments were performed under the approval of the Animal Care and Use Committee of Southwest Medical University.

References

1. L.G. Cooney, A. Dokras, Beyond fertility: Polycystic ovary syndrome and long-term health. Fertil. Steril. **110**, 794–809 (2018)
2. O. Osibogun, O. Ogunmoroti, E.D. Michos, Polycystic ovary syndrome and cardiometabolic risk: Opportunities for cardiovascular disease prevention. Trends Cardiovascular Med. **30**, 399–404 (2020)
3. E.M. Heijnen, M.J. Eijkemans, E.G. Hughes, J.S. Laven, N.S. Macklon, B.C. Fauser, A meta-analysis of outcomes of conventional IVF in women with polycystic ovary syndrome. Hum. Reprod. Update **12**, 13–21 (2006)
4. M.L. Pfieffer, Polycystic ovary syndrome: An update. Nursing **49**, 34–40 (2019)
5. S.U.A. Rasool, S. Ashraf, M. Nabi, F. Rashid, K.M. Fazili, S. Amin, Elevated fasting insulin is associated with cardiovascular and metabolic risk in women with polycystic ovary syndrome. Diabetes Metab. Syndr. **13**, 2098–2105 (2019)
6. R.P. Crespo, T. Bachega, B.B. Mendonça, L.G. Gomes, An update of genetic basis of PCOS pathogenesis. Arch. Endocrinol. Metab. **62**, 352–361 (2018)
7. E. Stener-Victorin, Q. Deng, Epigenetic inheritance of polycystic ovary syndrome - challenges and opportunities for treatment. Nat. Rev. Endocrinol. **17**, 521–533 (2021)
8. D.H. Abbott, M. Kraynak, D.A. Dumesic, J.E. Levine, In utero androgen excess: A developmental commonality preceding polycystic ovary syndrome? Front. Horm. Res. **53**, 1–17 (2019)

9. D. Dewailly, [Physiopathology of polycystic ovary syndrome]. Annales d'endocrinologie 60, 123–130 (1999)

10. G. Pei, L. Chen, W. Zhang, WGCNA application to proteomic and metabolomic data analysis. Methods Enzymol. 585, 135–158 (2017)

11. B. Zhang, S. Horvath, A general framework for weighted gene co-expression network analysis. Stat. Appl. Genet. Mol. Biol. 4, Article17 (2005)

12. P. Langfelder, S. Horvath, WGCNA: An R package for weighted correlation network analysis. BMC Bioinforma. 9, 559 (2008)

13. J.S. Amberger, C.A. Bocchini, F. Schiettecatte, A.F. Scott, A. Hamosh, OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. Nucleic acids Res. 43, D789–798 (2015)

14. S. Federhen, The NCBI taxonomy database. Nucleic acids Res. 40, D136–143 (2012)

15. Z. Mao, T. Li, H. Zhao, Y. Qin, X. Wang, Y. Kang, Identification of epigenetic interactions between microRNA and DNA methylation associated with polycystic ovarian syndrome. J. Hum. Genet. 66, 123–137 (2021)

16. J. Li, H. Chen, M. Gou, C. Tian, H. Wang, X. Song, D.L. Keefe, X. Bai, L. Liu, Molecular features of polycystic ovary syndrome revealed by transcriptome analysis of oocytes and cumulus cells. Front. Cell Developmental Biol. 9, 735684 (2021)

17. L. Jin, Q. Yang, C. Zhou, L. Liu, H. Wang, M. Hou, Y. Wu, F. Shi, J. Sheng, H. Huang, Profiles for long non-coding RNAs in ovarian granulosa cells from women with PCOS with or without hyperandrogenism. Reprod. biomedicine online 37, 613–623 (2018)

18. A. Alameer, D. Chicco. geoCancerPrognosticDatasetsRetriever, a bioinformatics tool to easily identify cancer prognostic datasets on Gene Expression Omnibus (GEO). (Bioinformatics, Oxford, England), 2021)

19. Gene Ontology Consortium: going forward. Nucleic Acids Res. 43, D1049–1056 (2015)

20. M. Kanehisa, S. Goto, KEGG: Kyoto encyclopedia of genes and genomes. Nucleic acids Res. 28, 27–30 (2000)

21. J. Piñero, J. Saüch, F. Sanz, L.I. Furlong, The DisGeNET cytoscape app: Exploring and visualizing disease genomics data. Computational Struct. Biotechnol. J. 19, 2960–2967 (2021)

22. J.D. Storey, R. Tibshirani, Statistical significance for genomewide studies. Proc. Natl. Acad. Sci. USA 100, 9440–9445 (2003)

23. D. Szklarczyk, A.L. Gable, D. Lyon, A. Junge, S. Wyder, J. Huerta-Cepas, M. Simonovic, N.T. Doncheva, J.H. Morris, P. Bork, L.J. Jensen, C.V. Mering, STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 47, D607–d613 (2019)

24. C.M. Labbé, J. Rey, D. Lagorce, M. Vavruša, J. Becot, O. Sperandio, B.O. Villoutreix, P. Tufféry, M.A. Miteva, MTiOpenScreen: A web server for structure-based virtual screening. Nucleic Acids Res. 43, W448–454 (2015)

25. O. Trott, A.J. Olson, AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J. Computational Chem. 31, 455–461 (2010)

26. H. Kemmish, M. Fasnacht, L. Yan, Fully automated antibody structure prediction using BIOVIA tools: Validation study. PloS one 12, e0177923 (2017)

27. E.F. Pettersen, T.D. Goddard, C.C. Huang, E.C. Meng, G.S. Couch, T.I. Croll, J.H. Morris, T.E. Ferrin, UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein science: a publication of the Protein. Society 30, 70–82 (2021)

28. A.M.S. Riel, D.A. Decato, J. Sun, C.J. Massena, M.J. Jessop, O.B. Berryman, The intramolecular hydrogen bonded-halogen bond: A new strategy for preorganization and enhanced binding. Chem. Sci. 9, 5828–5836 (2018)

29. Y.L. Yang, L.F. Sun, Y. Yu, T.X. Xiao, B.B. Wang, P.G. Ren, H.R. Tang, J.V. Zhang, Deficiency of Gpr1 improves steroid hormone abnormality in hyperandrogenized mice. Reprod. Biol. Endocrinol.: RBE 16, 50 (2018)

30. A.C. Ozdemir, G.M. Wynn, A. Vester, M.N. Weitzmann, G.N. Neigh, S. Srinivasan, M.K. Rudd, GNB3 overexpression causes obesity and metabolic syndrome. PloS one 12, e0188763 (2017)

31. C. Xu, J. Min, Structure and function of WD40 domain proteins. Protein cell 2, 202–214 (2011)

32. M. Schapira, M. Tyers, M. Torrent, C.H. Arrowsmith, WD40 repeat domain proteins: a novel target class? Nat. Rev. Drug Discov. 16, 773–786 (2017)

33. B.P. Jain, S. Pandey, WD40 Repeat Proteins: Signalling scaffold with diverse functions. Protein J. 37, 391–406 (2018)

34. V. Migliori, M. Mapelli, E. Guccione, On WD40 proteins: Propelling our knowledge of transcriptional control? Epigenetics 7, 815–822 (2012)

35. M.A. Villamil, Q. Liang, Z. Zhuang, The WD40-repeat protein-containing deubiquitinase complex: catalysis, regulation, and potential for therapeutic intervention. Cell Biochem. Biophys. 67, 111–126 (2013)

36. H. Tummala, S. Fleming, P.M. Hocking, D. Wehner, Z. Naseem, M. Ali, C.F. Inglehearn, N. Zhelev, D.H. Lester, The D153del mutation in GNB3 gene causes tissue specific signalling patterns and an abnormal renal morphology in Rge chickens. PloS one 6, e21156 (2011)

37. S. Klenke, M. Kussmann, W. Siffert, The GNB3 C825T polymorphism as a pharmacogenetic marker in the treatment of hypertension, obesity, and depression. Pharmacogenetics genomics 21, 594–606 (2011)

38. K.D. Ko, K.K. Kim, H.S. Suh, I.C. Hwang, Associations between the GNB3 C825T polymorphism and obesity-related metabolic risk factors in Korean obese women. J. endocrinological Investig. 37, 1117–1120 (2014)

39. H.F. Escobar-Morreale, Polycystic ovary syndrome: Definition, aetiology, diagnosis and treatment. Nat. Rev. Endocrinol. 14, 270–284 (2018)

40. W.A. March, V.M. Moore, K.J. Willson, D.I. Phillips, R.J. Norman, M.J. Davies, The prevalence of polycystic ovary syndrome in a community sample assessed under contrasting diagnostic criteria. Hum. Reprod. (Oxf., Engl.) 25, 544–551 (2010)

41. Y. Zhu, X. Lin, X. Zong, S. Han, M. Wang, Y. Su, L. Ma, X. Chu, C. Yi, Q. Zhao, B. Wu, Structural basis of FPR2 in recognition of Aβ(42) and neuroprotection by humanin. Nat. Commun. 13, 1775 (2022)

42. Z. Wu, Y. Wen, G. Fan, H. He, S. Zhou, L. Chen, HEMGN and SLC2A1 might be potential diagnostic biomarkers of steroid-induced osteonecrosis of femoral head: study based on WGCNA and DEGs screening. BMC Musculoskelet. Disord. 22, 85 (2021)

Affiliations

Gan Qiao[1] · Jinshan Xing[2] · Xin Luo[1] · Chunxiang Zhang[3] · Jingyan Yi [4]

[1] Department of Pharmacology, School of Pharmacy, Nucleic Acid Medicine of Luzhou Key Laboratory, Central Nervous System Drug Key Laboratory of Sichuan Province, Southwest Medical University, Luzhou 646000 Sichuan, China

[2] Department of Neurosurgery, The Affiliated Traditional Chinese Medicine Hospital of Southwest Medical University, Luzhou 646000 Sichuan, China

[3] Nucleic Acid Medicine of Luzhou Key Laboratory, Key Laboratory of Medical Electrophysiology, Ministry of Education & Medical Electrophysiological Key Laboratory of Sichuan Province, (Collaborative Innovation Center for Prevention of Cardiovascular Diseases), Institute of Cardiovascular Research, Southwest Medical University, Luzhou 646000 Sichuan, China

[4] Department of Medical Cell Biology and Genetics, School of Basic Medical Sciences, Nucleic Acid Medicine of Luzhou Key Laboratory, Key Laboratory of Medical Electrophysiology, Ministry of Education & Medical Electrophysiological Key Laboratory of Sichuan Province, (Collaborative Innovation Center for Prevention of Cardiovascular Diseases), Institute of Cardiovascular Research, Southwest Medical University, Luzhou 646000 Sichuan, China