Identification of Crucial lncRNAs, miRNAs, mRNAs, and Potential Therapeutic Compounds for Polycystic Ovary Syndrome by Bioinformatics Analysis

Zhi Zeng, Xia Lin, Tingting Xia, Wenxiu Liu, Xiaohui Tian, and Manchao Li

1Center of Reproductive Medicine, The Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou 510655, China
2Department of Gynecology, The Third Affiliated Hospital of Sun Yat-sen University, Guangzhou 510630, China
3Center for Reproductive Medicine, The Third Affiliated Hospital of Sun Yat-sen University, Guangzhou 510630, China
4Department of Gynecology, The First Affiliated Hospital of Guangzhou University of Chinese Medicine, Guangzhou 510407, China
5Department of Obstetrics and Gynecology, The Seventh Affiliated Hospital of Sun Yat-sen University, Shenzhen 518107, China

Correspondence should be addressed to Xiaohui Tian; tianxhui@mail.sysu.edu.cn and Manchao Li; limch3@mail.sysu.edu.cn

Zhi Zeng and Xia Lin contributed equally to this work.

Received 15 July 2020; Revised 23 September 2020; Accepted 3 October 2020; Published 6 November 2020

Academic Editor: Nadia M. Hamdy

Copyright © 2020 Zhi Zeng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Background. This study was aimed at mining crucial long noncoding RNAs (lncRNAs), microRNAs (miRNAs), and messenger RNAs (mRNAs) for the development of polycystic ovary syndrome (PCOS) based on the coexpression and the competitive endogenous RNA (ceRNA) theories and investigating the underlying therapeutic drugs that may function by reversing the expression of lncRNAs, miRNAs, and mRNAs.

Methods. RNA (GSE106724, GSE114419, GSE137684, and GSE138518) or miRNA (GSE84376 and GSE138572) expression profile datasets of PCOS patients were downloaded from the Gene Expression Omnibus database. The weighted gene coexpression network analysis (WGCNA) using four RNA datasets was conducted to construct the lncRNA-mRNA coexpression networks, while the common differentially expressed miRNAs in two miRNA datasets and module RNAs were used to establish the ceRNA network. A protein-protein interaction (PPI) network was created to explore the potential interactions between genes. Gene Ontology and KEGG pathway enrichment analyses were performed to explore the functions of genes in networks. Connectivity Map (CMap) and Comparative Toxicogenomics Database (CTD) analyses were performed to identify potential therapeutic agents for PCOS. Results. Three modules (black, magenta, and yellow) were identified to be PCOS-related after WGCNA analysis, in which KLF3-AS1-PLCG2, MAPKAPK5-AS1-MAP3K14, and WW22-AS2-TXNIP were important coexpression relationship pairs. WW22-AS2-hsa-miR-382-PLCG2 was a crucial ceRNA loop in the ceRNA network. The PPI network showed that MAP3K14 and TXNIP could interact with hub genes PLK1 (degree = 21) and TLR1 (degree = 18), respectively. These genes were enriched into mitosis (PLK1), immune response (PLCG2 and TLR1), and cell cycle (TXNIP and PLK1) biological processes. Ten small molecule drugs (especially quercetin) were considered to be therapeutical for PCOS.

Conclusion. Our study may provide a novel insight into the mechanisms and therapy for PCOS.

1. Introduction

Polycystic ovary syndrome (PCOS), characterized by polycystic ovary morphology, hyperandrogenism, and ovulatory disturbance, is the most frequent endocrine disorder in women of the reproductive age, affecting approximately 6%-9% of women worldwide [1]. PCOS is not only a leading cause of female infertility [2] but also an important risk factor for insulin resistance, diabetes, obesity, hypertension, metabolic syndrome, cardiovascular disease, and endometrial cancer [3]. However, the treatment of PCOS remains a challenge; thus, it is of significance to investigate the
mechanisms and develop more effective preventative and therapeutic strategies.

Although the pathogenesis of PCOS has not been fully understood, apoptosis of ovarian granulosa cells (OGCs) is considered to be a potential contributor [4, 5]. By binding to the receptor on the OGCs, a follicle-stimulating hormone induces the expression of cholesterol side-chain cleavage cytochrome P450 (CYP19A1) and 17β-hydroxysteroid dehydrogenase (CYP17A1) which are rate-limiting enzymes to catalyze conversion from estrogen to estradiol [6]. Apoptosis of OGCs leads to the deficiency of estradiol and blocks the follicle development and maturation, ultimately forming more numbers of hypoprogenetic secondary follicles undergoing atresia and antral follicles [7]. Therefore, exploration of molecular mechanisms associated with the apoptosis of OGCs may provide underlying targets for developing drugs.

Accumulating evidence has suggested that long non-coding RNAs (lncRNAs), microRNAs (miRNAs), and messenger RNA (mRNAs) are important regulators of cell physiological and pathological processes (such as proliferation and apoptosis): mRNAs that encode the proteins were directly functional; miRNAs, approximately 22 nucleotide-long noncoding RNAs, act by binding to complementary sequences in the 3’ untranslated region (UTR) of mRNAs and then to negatively regulate the expressions of these mRNAs [8]; lncRNAs, >200 bp in length, can interact with the mRNA as a competing endogenous RNA (ceRNA) to regulate the expression of target genes [9] or directly regulate the miRNA as a competing endogenous RNA (ceRNA) to catalyze conversion from estrogen to estradiol [6]. Apoptosis of OGCs and Ovarian granulosa-like tumor cells line KGN, and significantly promoted apoptosis of ovarian granulosa-like tumor cell line KGN, and significantly enhanced the expression of proapoptotic Bax and p53 [5]. Li et al. reported that lncRNA SRLR was upregulated in PCOS patients compared with healthy females. Overexpression of lncRNA SRLR promoted apoptosis of KGN cells by upregulation of interleukin-6 (IL-6) [11]. The study of Fu et al. revealed that miR-16 expression was downregulated in ovarian cancer tissues of PCOS patients. Overexpression of miR-16 inhibited apoptosis of OGC in vitro and alleviated PCOS in vivo, which was associated with its function to downregulate its direct target programmed cell death 4 (PDCD4). Enforced expression of PDCD4 reversed the effects of miR-16 on OGC apoptosis [12]. Liu et al. demonstrated that lncRNA PVT1 may regulate the progression of PCOS by modulating miR-17-5p-PTEN: overexpressed PVT1 could bind with and inhibit the expression of miR-17-5p in OGCs of PCOS, thereby preventing the inhibitory effects of miR-17-5p for its target gene PTEN and leading to the elevated expression of PTEN. shRNA-mediated silencing of PVT1 and transfection with miR-17-5p mimics could decrease apoptosis, while accelerate colony formation ability and proliferation of OGCs [13].

Wang et al. observed that cotreatment with metformin and sitagliptin attenuated the apoptosis in PCOS model cells by inducing lncRNA H19 expression and proposed the three-combined strategy (metformin-sitagliptin-H19-expressing lentiviruses) for the treatment of PCOS [14]. However, crucial lncRNAs, miRNAs, mRNAs, and drugs that target these molecules remain rarely reported for PCOS.

Recently, with the development of the sequencing technique, there were some studies to mine the key lncRNAs [15], miRNAs [16–18], and mRNAs [15, 17, 19] in OGCs of PCOS patients using the gene or miRNA expression profile data. However, all of them focused only on the differentially expressed mRNAs (coding genes) (DEGs) [15, 19], lncRNAs (DEls) [15], miRNAs (DEMs) alone [16, 18], or DEM-DEG [17]. Also, the small molecule drugs that targeted them were not investigated. In this study, we aimed to screen pivotal lncRNAs, miRNAs, and mRNAs based on lncRNA-mRNA coexpression and lncRNA-miRNA-miRNA ceRNA mechanisms via integrated analysis of multiple high-throughput datasets. In addition, the Connectivity Map (CMap) and Comparative Toxicogenomics Database (CTD) analyses were performed to acquire bioactive compounds that may have potential for the treatment of PCOS by regulating the expression of DEGs, DELs, and DEMs.

2. Materials and Methods

2.1. Microarray Data. Multiple high-throughput datasets of PCOS were downloaded from the Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) database on February 25, 2020: (a) mRNA+lncRNA expression profile, including (1) GSE106724: it contained 12 samples of ovarian granulosa cells from normal control (n = 4) and PCOS (n = 8). Its analysis platform was Agilent-062918 Human IncRNA array V4.0 (Probe Name version). (2) GSE114419: 6 subjects [normal control (n = 3) and PCOS (n = 3)] were enrolled for obtaining their ovarian granulosa cells. Its analysis platform was Affymetrix Human Transcriptome Array 2.0 [transcript (gene) version] [20]. (3) GSE137684: ovarian granulosa cells of 4 normal control and 8 PCOS patients were available. Its analysis platform was Agilent-039494 SurePrint G3 Human GE v2 8x60K Microarray 039381 (Probe Name version). (4) GSE138518: RNA sequencing in ovarian granulosa cells of PCOS patients (n = 3) and normal people (n = 3), which were run on an Illumina HiSeq 2000 platform. (b) miRNA expression profile, including (1) GSE138572: small RNA sequencing by an Illumina HiSeq 2000 platform was applied in ovarian granulosa cells of PCOS patients (n = 5) and normal people (n = 5) and (2) GSE84376: granulosa cells of 15 PCOS patients and 13 control individuals were obtained. Its miRNA analysis was carried out using the Affymetrix GeneChip miRNA 3.0 array.

2.2. Differential Expression Analysis. The mRNAs and lncRNAs in the above four datasets were reannotated according to the official nomenclature in the HUGO Gene Nomenclature Committee (HGNC; http://www.genenames.org/) [21]. Only the shared mRNAs and lncRNAs in all of the included datasets were obtained for the differential
analysis. The DELs and DEGs between PCOS and normal controls were identified using the MetaDEES function in the MetaDE package (version 1.0.5, https://cran.r-project.org/web/packages/MetaDE/). The threshold was set based on three aspects: (1) no heterogeneity was present among these datasets, which was assessed by tau^2 statistic (=0) and chi-square-based Q-test (p > 0.05); (2) false discovery rate (FDR) was less than 0.05; and (3) the differential trend was consistent based on log2FC (fold change).

The Linear Models for Microarray Data (LIMMA) package (version 3.34.7; https://bioconductor.org/packages/release/bioc/html/limma.html) [22] for R was used for identification of DEMs in GSE138572 and GSE84376 datasets, with [logFC] > 0.263 and false discovery rate (FDR) < 0.05 defined as the cut-off point. The DEMs commonly screened in two datasets were shown by Venn diagram. The volcano plot was visualized using the ggplot2 package (version 3.3.0; https://cran.r-project.org/web/packages/ggplot2) in R. A heat map was created using the "pheatmap" package (version: 1.0.8; https://cran.r-project.org/web/packages/pheatmap) in R.

2.3. Crucial IncRNA and mRNA Identification for PCOS. Weighted gene coexpression network analysis (WGCNA) package in R (version 1.61; https://cran.r-project.org/web/packages/WGCNA/index.html) [23] was used to cluster mRNAs and lncRNAs into several coexpression modules in which genes were highly correlated and considered to be crucial. During this WGCNA analysis, four steps were included: (1) calculating the expression and connectivity correlations of IncRNAs and mRNAs between any two datasets (GSE106724, GSE114419, GSE137684, and GSE138518), with correlation coefficient > 0.6 and p value > 0.001 defined as significantly correlated; (2) using the GSE106724 dataset to select the soft threshold power (β) based on the scale-free topology criterion (that is, when the R^2 reached 0.9 for the first time); (3) generating the dendrogram by calculation of the topological overlap matrix dissimilarity between genes in the GSE106724 dataset and identifying gene modules (cutHeight = 0.99 and minSize ≥ 100) by the dynamicTreeCut method [24]; (4) assessing the preservation (Z-score > 5 and p < 0.05) of the identified modules in the GSE114419, GSE137684, and GSE138518 datasets using the modulePreservation statistics [25]; (5) mapping the DEGs and DELs into the modules and screening the modules with more DEGs and DELs (p < 0.05, fold enrichment > 1) by using the hypergeometric algorithm [f(k, N, M, n) = C(k, M) * C(n − k, N − M)/C(n, N)] [26]; and (6) correlating coexpression genes modules with PCOS development by moduleTraitCor and moduleTraitPvalue algorithms.

2.4. IncRNA-miRNA Coexpression Network Visualization. The associations between the DEGs and DELs screened in the crucial WGCNA modules were evaluated by using the cor.test function (https://stat.ethz.ch/R-manual/R-devel/library/stats/html/cor.test.html) in R, and then the coexpression network was constructed based on their Pearson correlation coefficients (PCC). The coexpression network was visualized using Cytoscape (version 3.6.1; http://www.cytoscape.org/).

2.5. IncRNA-miRNA-mRNA ceRNA Network Construction. Interaction relationships of module DELs and DEGs with miRNAs were, respectively, predicted with the DIANA-LncBase (v2.0; http://carolina.imis.athena-innovation.gr/diana_tools/web/index.php?r=Lncbasev2/index-predicted) [27] and starBase (v2.0; http://starbase.sysu.edu.cn/) [28] databases. IncRNA-miRNA-mRNA interaction axes were selected according to the overlapped miRNAs interacted with DEGs and DELs. The Cytoscape (version 3.6.1; http://www.cytoscape.org/) software was used for visualization of the ceRNA regulatory network.

2.6. Protein-Protein Interaction (PPI) Network Construction. The interactions between proteins coding by the module DEGs were predicted using the STRING (Search Tool for the Retrieval of Interacting Genes; version 10.0; https://string-db.org) database [29]. PPIs with a confidence score ≥ 0.4 were chosen to construct the network. The PPI network was visualized by Cytoscape software (version 3.6.1; http://www.cytoscape.org/). Furthermore, the CytoNCA plugin in Cytoscape software (http://apps.cytoscape.org/apps/cytonica) [30] was applied to explore the hub genes by calculating the topological features of each protein in the PPI network, including degree centrality (DC), betweenness centrality (BC), and closeness centrality (CC). The proteins with high DC, BC, and CC were considered as hubs and essential for PCOS.

2.7. Function Enrichment Analysis. The Gene Ontology (GO) annotation, Kyoto Encyclopedia of Genes and Genomes (KEGG), and REACTOME pathway analyses were conducted for the DEGs in the PPI network using the online Database for Annotation, Visualization and Integrated Discovery (DAVID) (version 6.8; http://david.abcc.ncifcrf.gov) [31], p value < 0.05 under hypergeometric test was considered as statistically significant.

2.8. Small Molecule Drug Analysis. The CMap (https://portals.broadinstitute.org/cmap/) was used to identify candidate small molecule drugs that may have potential to treat PCOS patients. The identified module DEGs were uploaded into the CMap web tool, and then the resultant small molecules with negative connectivity scores were considered to be therapeutic agents because it could reverse the expression of the query signature. The small molecules with the enrichment score approximate to -1 may be especially effective. Furthermore, CTD (http://ctdbase.org) was also searched using the gene name, after which a serial of chemical-gene interaction pairs were obtained. The small molecules identified by CMap and CTD were then compared with these chemical-gene interaction pairs to screen the common. The gene-drug interaction networks were visualized using Cytoscape (version 3.6.1; http://www.cytoscape.org/).

3. Results

3.1. Identification of DELs, DEGs, and DEMs. A total of 70 IncRNAs and 10,481 protein-encoding mRNAs were annotated in all GSE106724, GSE114419, GSE137684, and GSE138518 datasets, which then underwent MetaDE analysis...
Figure 1: Continued.
to identify DELs and DEGs. As a result, 754 differentially expressed RNAs were found between PCOS patients and normal controls, including 19 DELs (upregulated, 18; downregulated, 1) and 735 DEGs (upregulated, 551; downregulated, 184) (Table S1). Heat map analysis showed that these DELs and DEGs can obviously cluster the samples into two groups in any one dataset (Figure 1(a)).

Furthermore, LIMMA analysis was used to screen DEMs in GSE138572 and GSE84376, respectively. The results showed that 80 DEMs (upregulated, 16; downregulated, 64) were identified for the GSE138572 dataset (Figure 1(b)) (Table S2), while 106 (upregulated, 74; downregulated, 32) for the GSE84376 dataset (Figure 1(d)) (Table S3). These DEMs also can obviously distinguish the PCOS patients from normal controls in the GSE138572 (Figure 1(c)) and GSE84376 (Figure 1(e)) datasets. The Venn diagram revealed that there were 11 DEMs (including 4 upregulated and 7 downregulated) consistently expressed in these two datasets (Figure 1(f)).

3.2. Construction of IncRNA-mRNA Coexpression Network Using Module Genes. Based on GSE106724, GSE114419, GSE137684, and GSE138518 datasets, WGCNA was applied to detect the PCOS-related modules which included the potential interactions between IncRNAs and mRNAs. These four datasets were considered to be comparable because there were significantly positive correlations of RNAs (coefficient > 0.6 and p value < 1e-200) between any two datasets irrespective of their expression levels (Figure 2(a)) or the connectivity (Figure 2(b)). As shown in Figures 2(c) and 2(d), power = 20 was chosen as the soft-thresholding to ensure a scale-free network ($R^2 = 0.9$; mean connectivity = 1). A cluster dendrograms showed that eleven colors of modules (black, blue, brown, green, grey, magenta, pink, purple, red, turquoise, and yellow, ranging in size from 136 to 2,812) were identified using the GSE106724 dataset (Figure 3(a); Table 1). These modules were also observed in the analysis of GSE138518, GSE137684, and GSE114419 datasets (Figures 3(b)–3(d)). Among these eleven modules, black, blue, magenta, purple, turquoise, and yellow may be preserved because of their $Z$-score > 5 and p value < 0.05 (Table 1). Black, magenta, and yellow modules were further considered to be crucial for PCOS because there were relatively many DEGs or DELs enriched in them (enrichment fold > 1 and p value < 0.05) (Table 1). This conclusion was also demonstrated according to the
Figure 2: Continued.


correlation coefficients and their p-values for different datasets.
results of a correlation heat map in which black ($p = 1e-24$), magenta ($p = 0$), and yellow ($p = 1e-12$) modules were significantly, positively related with the development of PCOS (Figure 3(e)). Subsequently, the expressions of DELs and DEGs in these three modules were extracted and the PCC were calculated to construct the coexpression network. As shown in Figure 4, all the included genes in the black module were DEGs (20) (Figure 4(a)), while the magenta module (Figure 4(b)) had 1 DEL (SMCR5) and 20 DEGs; the yellow module (Figure 4(c)) had 4 DELs (KLF3-AS1, LINC00910, MAPKAPK5-AS1, and WWC2-AS2) and 45 DEGs, indicating the 160 interaction pairs of the magenta and yellow modules may be especially pivotal such as LINC00910-thioredoxin interacting protein (TXNIP), KLF3-AS1-phospholipase C gamma 2 (PLCG2), MAPKAPK5-AS1-mitogen-activated protein kinase kinase kinase 14 (MAP3K14), and WWC2-AS2-TXNIP for PCOS.

3.3. Construction of the ceRNA Regulatory Network. The ceRNA regulatory network was also constructed using the DELs and DEGs in the black, magenta, and yellow modules. A total of 4 interaction relationship pairs between 4 DELs (LINC00910, MAPKAPK5-AS1, SMCR5, and WWC2-AS2) and 4 DEMs (hsa-miR-455, hsa-miR-4467, hsa-miR-4665, and hsa-miR-382) were predicted by the DIANA-LncBase database, while 12 paired interactions were predicted between 4 DEMs and 9 DEGs by the starBase database. Then, the lncRNA-miRNA-mRNA network was established based on these ceRNA loops (Figure 5), including SMCR5-hsa-miR-4665-G protein signaling modulator 2 (GSM2), LINC00910-hsa-miR-435-polo like kinase 1 (PLK1), MAPKAPK5-AS1-hsa-miR-4467-PLEKHG3, and WWC2-AS2-hsa-miR-382-PLCG2. However, only WWC2-AS2-hsa-miR-382-PLCG2 may be believable because the expressions of WWC2-AS2 (upregulated) and PLCG2 (upregulated) were opposite to hsa-miR-382 (downregulated).

3.4. Construction of the PPI Network. In order to provide the potential interpretations for the DEGs that may have no function enrichment, a PPI network was constructed for the DEGs in crucial modules. As a result, a total of 280 interaction pairs were identified, such as MAP3K14-PLK1 and TXNIP-toll-like receptor 1 (TLR1) (Figure 6). PLK1 and TLR1 were considered as hub genes in the PPI because they belonged to the top 20 genes ranked by DC, BC, and CC (Table 2). 3.5. Function Enrichment Analysis. A total of 18 GO biological process terms were enriched for the DEGs in the PPI network (Table 3), such as GO:0000280 - nuclear division (PLK1), GO:0007067 - mitosis (PLK1), GO:0006955 - immune response (PLCG2 and TLR1), and GO:0007049 - cell cycle (TXNIP and PLK1). Two KEGG pathways were also obtained (Table 3), including hsa05120: epithelial cell signaling in Helicobacter pylori infection (PLCG2 and MAP3K14). 3.6. Identification of Small Molecule Drugs. A total of 629 small molecules were found to have the negative enrichment score, indicating they may be potential therapeutical drugs for PCOS. Ten small molecule drugs enriched by CMap (irinotecan, score = −0.902; doxorubicin, score = −0.787; imatinib, score = −0.714; etoposide, score = −0.696; methotrexate, score = −0.6; promethazine, score = −0.563; gentamicin, score = −0.562; thapsigargin, score = −0.548; quercetin, score = −0.538; and valproic acid, score = −0.192) were identified to target crucial genes in CTD database analysis (Figure 7). Among them, valproic acid could target any genes in the ceRNA axes of WWC2-AS2-hsa-miR-382-PLCG2 and coexpression axes of KL3-AS1-PLCG2, MAPKAPK5-AS1-MAP3K14, and WWC2-AS2-TXNIP. Furthermore, doxorubicin (PLK1, PLCG2, TXNIP, and TLR1) and methotrexate (PLK1, TXNIP, MAP3K14, and TLR1) targeted 4 mRNAs; quercetin (PLK1, TXNIP, and MAP3K14) and thapsigargin.
Figure 3: WGCNA analysis. (a–d) Clustering of differentially expressed lncRNAs and mRNAs of GSE106724 (a), GSE138518 (b), GSE137684 (c), and GSE114419 (d) datasets; (e) the correlation between gene modules and PCOS development.
Table 1: Identified modules by WGCNA analysis.

<table>
<thead>
<tr>
<th>ID</th>
<th>Color</th>
<th>Module size</th>
<th>Preservation information</th>
<th>#DEGs</th>
<th>Enrichment information</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Z-score</td>
<td>p value</td>
<td></td>
</tr>
<tr>
<td>Module 1</td>
<td>Black</td>
<td>150</td>
<td>7.0208</td>
<td>2.0E-10</td>
<td>26</td>
</tr>
<tr>
<td>Module 2</td>
<td>Blue</td>
<td>332</td>
<td>10.5576</td>
<td>1.90E-03</td>
<td>10</td>
</tr>
<tr>
<td>Module 3</td>
<td>Brown</td>
<td>327</td>
<td>1.5476</td>
<td>1.90E-06</td>
<td>17</td>
</tr>
<tr>
<td>Module 4</td>
<td>Green</td>
<td>213</td>
<td>0.3623</td>
<td>8.60E-01</td>
<td>9</td>
</tr>
<tr>
<td>Module 5</td>
<td>Grey</td>
<td>2,812</td>
<td>0.9776</td>
<td>3.10E-01</td>
<td>259</td>
</tr>
<tr>
<td>Module 6</td>
<td>Magenta</td>
<td>137</td>
<td>5.1445</td>
<td>8.60E-06</td>
<td>21</td>
</tr>
<tr>
<td>Module 7</td>
<td>Pink</td>
<td>143</td>
<td>0.4311</td>
<td>5.10E-01</td>
<td>12</td>
</tr>
<tr>
<td>Module 8</td>
<td>Purple</td>
<td>136</td>
<td>7.9183</td>
<td>1.40E-06</td>
<td>1</td>
</tr>
<tr>
<td>Module 9</td>
<td>Red</td>
<td>152</td>
<td>0.7328</td>
<td>8.20E-01</td>
<td>9</td>
</tr>
<tr>
<td>Module 10</td>
<td>Turquoise</td>
<td>519</td>
<td>6.5919</td>
<td>9.80E-04</td>
<td>43</td>
</tr>
<tr>
<td>Module 11</td>
<td>Yellow</td>
<td>245</td>
<td>19.8755</td>
<td>2.10E-11</td>
<td>45</td>
</tr>
</tbody>
</table>

DEGs: differentially expressed genes; CI: confidence interval.

(PLK1, PLCG2, and TXNIP) targeted 3 mRNAs; etoposide and irinotecan targeted 2 mRNAs (PLK1 and TXNIP); and imatinib (PLK1), gentamicin (TXNIP), and promethazine (PLK1) targeted 1 mRNA.

4. Discussion

Although previous studies have attempted to reveal the molecular mechanisms of PCOS by using high-throughput microarray or sequencing data [15–19], all of them focused on the single dataset (mRNA: GSE95728 [15], miRNA: GSE84376 [16, 18], and mRNA: GSE34526 [17]) which may result in a high probability of false positive results. In order to prevent this shortcoming, in this study, we analyzed four lncRNA-mRNA expression profile datasets (GSE106724, GSE114419, GSE137684, and GSE138518) using the MetaDE package and only selected the DELs and DEGs that overlapped in the four datasets. The crucial DELs and DEGs were screened via WGCNA analysis which used GSE106724 as the training dataset and GSE114419, GSE137684, and GSE138518 as the validation datasets. Similarly, two miRNA expression profile datasets were used. Although some key miRNAs and mRNAs identified in the previous studies (miR-3135, miR-3188 [16, 18], and miR-486 [17]; aquaporin 9, free fatty acid receptor 2, and S100 calcium binding protein A8 [15]) were also found in our study, they were not the focus because they were only identified in one dataset (miRNA, GSE84376) or not included in preserved modules. More interestingly, we, for the first time, integrated the DEL-DEG and DEL-DEM-DEG data to construct the coexpression and ceRNA network. As expected, we identified three new coexpression relationships between 3 DELs and their target genes (KLF3-AS1-PLCG2, MAPKAPK5-AS1-MAP3K14, and WW2-AS2-TXNIP) and one ceRNA axis among WW2-AS2, hsa-miR-382, and PLCG2. These mRNAs were predicted to be involved in PCOS by influencing cell mitosis, cell cycle, and immune response. These three DELs (KLF3-AS1, MAP-KAPK5-AS1, and WW2-AS2), one DEM (hsa-miR-382), and five DEGs [including 3 in the coexpression or ceRNA axes (PLCG2, MAP3K14, and TXNIP)] and 2 interacted genes (PLK1 and TLR1) in the PPI network were predicted to be regulated by 10 small molecular drugs for the treatment of PCOS via CMap and CTD analyses.

Among all these DELs, DEMs, and DEGs, only TXNIP, TXNIP-interacted TLR1, and miR-382-5p were directly reported to be associated with PCOS. TXNIP expression was downregulated during the time of in vitro culture of the OGCs [32–34]. Upregulation of TXNIP in the OGCs resulted in the patients suffering PCOS [35]. Higher serum TXNIP was also indicated to be associated with impaired β cell function and insulin resistance in PCOS patients [36]. The expressions of toll-like receptor signaling pathway genes (including TLR1, TLR2, TLR4, TLR8, and CD14) were found to be significantly increased in OGC samples of PCOS patients compared with controls [19, 37]. Thus, TXNIP and TLR1 may be involved in PCOS by activating the inflammatory pathway and then inducing cell apoptosis as reported in other diseases. For example, Shan et al. proved that knockdown of TXNIP decreased the inflammatory damage in kidney tissues induced by 2,2',4,4'-tetrabromodiphenyl ether [38]. The study of Chen et al. revealed that hypoxia induced pancreatic β cell death by upregulating the reactive oxygen species- (ROS-) TXNIP-NLR family pyrin domain containing 3 (NLRP3) inflammasome axis [39]. Also, the traditional Chinese medicine Lycium barbarum polysaccharide was demonstrated to attenuate hepatocyte apoptosis induced by ethanol through inhibiting the TXNIP-NLRP3 inflammasome pathway [40]. By using the littermate model, Mohamed et al. verified that TXNIP deletion ameliorated the inflammatory response in high fat diet-induced nonalcoholic steatohepatitis via deactivation of the TLR2-NLRP3 inflammasome axis [41]. Furthermore, miR-382-5p was shown to be negatively correlated with a free androgen index [42]. The PCOS group had a significantly higher free androgen index than controls [43]. In line with these studies, we also identified...
Figure 4: Continued.
Figure 4: The coexpression relationships in three crucial modules: (a) black module; (b) magenta module; (c) yellow module. Genes in boxes indicate the lncRNAs; genes in circles indicate the mRNAs.

Figure 5: Construction of the ceRNA network using the lncRNAs, mRNAs in three crucial modules, and common miRNAs identified by GSE84376 and GSE138572 datasets. Genes in boxes indicate the lncRNAs; genes in circles indicate the mRNAs; genes in triangles indicate the miRNAs. Different colors for lncRNAs and mRNAs represent different modules.
that TXNIP and TLR1 were expressed higher, while miR-382-5p was expressed lower, in the OGCs of PCOS patients than in normal controls.

Furthermore, *in vitro* exposure of human luteinized mural granulosa cells to dibutyl phthalate was found to induce the high expression of PLK1 [44]. Gestational
exposure to dibutyl phthalate was revealed to induce polycystic ovaries and a hormonal profile similar to PCOS [45]. Mass spectrometry analysis showed that culture of OGCs with 10 ng/mL fibroblast growth factor- (FGF-) 8 and FGF-18 significantly triggered upregulation of several proteins, including PLCG2 [46]. Increased intrafollicular FGF-13 levels were positively correlated with elevated total testosterone and increased ovarian volume, but negatively associated with homeostatic model assessment of insulin resistance of PCOS patients [47]. These findings indirectly demonstrated their possible roles in PCOS. In accordance with these studies, we also identified that PLK1 and PLCG2 were upregulated in the OGCs of PCOS patients compared with the placebo group [48]. For example, Khorshidi et al. observed that supplementing 1,000 mg/day quercetin for 12 weeks significantly decreased testosterone, fasting blood glucose, insulin, and inflammatory cytokines (IL-1β, IL-6, and tumor necrosis factor α) [49], increasing the activity of antioxidant enzymes (superoxide dismutase, catalase, glutathione-S-transferase, and reduced glutathione) and improving granulosa cell apoptosis (upregulation of Bcl-2 and downregulation of Bax) [50]. In this study, we predicted that quercetin targeted PLK1, TXNIP, and MAP3K14. Thus, correlated with miR-382. These findings indirectly explain the possible regulatory relationships between them.

More importantly, valproic acid, doxorubicin, methotrexate, quercetin, thapsigargin, etoposide, irinotecan, imatinib, gentamicin, and promethazine were predicted to treat PCOS by reversing the expressions of the above crucial DELs, DEMs, and DEGs in this study. Some of them were confirmed to have therapeutic potential for PCOS previously. For example, Khorshidi et al. observed that supplementation of in vivo quercetin for 12 weeks significantly decreased testosterone, fasting blood glucose, insulin, and homeostatic model assessment of insulin resistance of PCOS patients compared with the placebo group [48]. Quercetin alleviated the PCOS by suppressing the levels of inflammatory cytokines (IL-1β, IL-6, and tumor necrosis factor α) [49], increasing the activity of antioxidant enzymes (superoxide dismutase, catalase, glutathione-S-transferase, and reduced glutathione) and improving granulosa cell apoptosis (upregulation of Bcl-2 and downregulation of Bax) [50]. In this study, we predicted that quercetin targeted PLK1, TXNIP, and MAP3K14. Thus,

<table>
<thead>
<tr>
<th>Category</th>
<th>Term</th>
<th>p value</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>GOTERM_BP_FAT</td>
<td>GO:0000280 – nuclear division</td>
<td>2.58E-04</td>
<td>KIF11, PLK1, KIF15, BUB1, NDC80, AURKB, CDCA5</td>
</tr>
<tr>
<td>GOTERM_BP_FAT</td>
<td>GO:00007670 – mitosis</td>
<td>2.58E-04</td>
<td>KIF11, PLK1, KIF15, BUB1, NDC80, AURKB, CDCA5</td>
</tr>
<tr>
<td>GOTERM_BP_FAT</td>
<td>GO:00000087 – M phase of mitotic cell cycle</td>
<td>2.84E-04</td>
<td>KIF11, PLK1, KIF15, BUB1, NDC80, AURKB, CDCA5</td>
</tr>
<tr>
<td>GOTERM_BP_FAT</td>
<td>GO:0048285 – organelle fission</td>
<td>3.20E-04</td>
<td>KIF11, PLK1, KIF15, BUB1, NDC80, AURKB, CDCA5</td>
</tr>
<tr>
<td>GOTERM_BP_FAT</td>
<td>GO:0000278 – mitotic cell cycle</td>
<td>7.09E-04</td>
<td>KIF11, PLK1, KIF15, BUB1, NDC80, AURKB, CDCA5, GTSE1</td>
</tr>
<tr>
<td>GOTERM_BP_FAT</td>
<td>GO:0001775 – cell activation</td>
<td>1.05E-03</td>
<td>PREX1, PLCG2, TLRI, IRF1, NFAM1, LTB, ICOSLG</td>
</tr>
<tr>
<td>GOTERM_BP_FAT</td>
<td>GO:0022403 – cell cycle phase</td>
<td>1.36E-03</td>
<td>KIF11, PLK1, KIF15, BUB1, NDC80, AURKB, CDCA5, GTSE1</td>
</tr>
<tr>
<td>GOTERM_BP_FAT</td>
<td>GO:0006955 – immune response</td>
<td>1.71E-03</td>
<td>OSM, PO2AF1, CST7, ENPP3, PLCG2, TLRI, CD300LF, CCL5, LTG, ICOSLG</td>
</tr>
<tr>
<td>GOTERM_BP_FAT</td>
<td>GO:0000279 – M phase</td>
<td>2.10E-03</td>
<td>KIF11, PLK1, KIF15, BUB1, NDC80, AURKB, CDCA5</td>
</tr>
<tr>
<td>GOTERM_BP_FAT</td>
<td>GO:0045321 – leukocyte activation</td>
<td>2.95E-03</td>
<td>PREX1, PLCG2, TLRI, IRF1, NFAM1, ICOSLG</td>
</tr>
<tr>
<td>GOTERM_BP_FAT</td>
<td>GO:0051056 – regulation of small GTPase-mediated signal transduction</td>
<td>3.51E-03</td>
<td>PLEKHG3, TBC1D10C, GMIP, PREX1, RASAL3, ARHGAP27</td>
</tr>
<tr>
<td>GOTERM_BP_FAT</td>
<td>GO:0007017 – microtubule-based process</td>
<td>3.57E-03</td>
<td>KIF4A, KIF11, KIF15, NDC80, GTSE1, KIF20A</td>
</tr>
<tr>
<td>GOTERM_BP_FAT</td>
<td>GO:0051301 – cell division</td>
<td>6.81E-03</td>
<td>KIF11, PLK1, BUB1, NDC80, AURKB, CDCA5, GTSE1</td>
</tr>
<tr>
<td>GOTERM_BP_FAT</td>
<td>GO:0022402 – cell cycle process</td>
<td>7.62E-03</td>
<td>KIF11, PLK1, KIF15, BUB1, NDC80, AURKB, CDCA5, GTSE1</td>
</tr>
<tr>
<td>GOTERM_BP_FAT</td>
<td>GO:0046578 – regulation of Ras protein signal transduction</td>
<td>1.04E-02</td>
<td>PLEKH3G, TBC1D10C, GMIP, PREX1, ARHGAP27</td>
</tr>
<tr>
<td>GOTERM_BP_FAT</td>
<td>GO:0007049 – cell cycle</td>
<td>1.26E-02</td>
<td>TXNIP, KIF11, PLK1, KIF15, BUB1, NDC80, AURKB, CDCA5, GTSE1</td>
</tr>
<tr>
<td>GOTERM_BP_FAT</td>
<td>GO:0007242 – intracellular signaling cascade</td>
<td>2.84E-02</td>
<td>OSM, GMIP, PREX1, PLCG2, TLRI, PPM1L, NDC80, NFAM1, GRAPI, LTG, GTSE1</td>
</tr>
<tr>
<td>GOTERM_BP_FAT</td>
<td>GO:0006952 – defense response</td>
<td>3.79E-02</td>
<td>AOX1, TLRI, MND, NFAM1, CCL5, MX2, ICOSLG</td>
</tr>
<tr>
<td>KEGG_pathway</td>
<td>hsa00620: nicotinate and nicotinamide metabolism</td>
<td>5.08E-03</td>
<td>ENPP3, AOX1, QPRT</td>
</tr>
<tr>
<td>KEGG_pathway</td>
<td>hsa05120: epithelial cell signaling in Helicobacter pylori infection</td>
<td>3.72E-02</td>
<td>PLCG2, MAP3K14, CCL5</td>
</tr>
</tbody>
</table>

DEGs: differentially expressed genes; PPI: protein-protein interaction; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes.
we recommended combining quercetin supplementation and siRNA knockdown of PLK1, TXNIP, and MAP3K14 for the treatment of PCOS, which may be more effective than quercetin-alone treatment.

Some limitations of our study should be acknowledged. First, all of the lncRNAs were not previously confirmed to be associated with PCOS by wet experiments. Clinical samples should be collected to further validate their expressions and associations with androgen, MII oocyte rate, and infertility rate. In vitro and in vivo studies should be performed to explore their influence on apoptosis and proliferation of OGCs. Second, the coexpression relationship between lncRNAs and mRNAs and the ceRNA axes among lncRNAs, miRNAs, and mRNAs should be verified by immunoprecipitation, cooverexpression, or coknockdown experiments. Third, the target mechanisms of small molecular drugs on lncRNAs, miRNAs, and mRNAs and the effects of their combination therapy also needed further investigation. Fourth, the significantly differential lncRNAs, miRNAs, and mRNAs that are not commonly expressed in multiple datasets or enriched in WGCNA modules also should be given attention to confirm their importance.

5. Conclusion

Based on coexpression and ceRNA network analyses, the present study identified that several lncRNAs (KLF3-AS1, WWC2-AS, and MAPKAPK5-AS1), miRNAs (miR-382), and DEGs (PLK1, PLCG2, TXNIP, TLR1, and MAP3K14) were associated with the development of PCOS. In addition, ten small molecular drugs (such as quercetin) were predicted to be therapeutic agents for PCOS by reversing the expressions of these crucial genes. Our study may provide a novel insight into the mechanisms and therapy for PCOS.

Data Availability

All data were downloaded from the Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) under GSE106724, GSE114419, GSE137684, GSE138518, GSE138572, and GSE84376 accession numbers.

Conflicts of Interest

The authors declare that they have no competing interests.

Authors’ Contributions

ZZ, XL, XHT, and MCL conceived of the study and participated in its design. ZZ and XL downloaded, processed the data, and performed the related bioinformatic analyses. TTX and WXL contributed to the interpretation of the results. ZZ and XL drafted the article. XHT and MCL revised the manuscript. All authors have read and approved the final version of the article.

Supplementary Materials

Supplementary 1. Table S1: all differential lncRNAs and miRNAs in four datasets.
Supplementary 2. Table S2: all differential miRNAs in GSE13572 dataset.
Supplementary 3. Table S3: all differential miRNAs in GSE84376 dataset.

References


