Medicine®

OPEN

Identification of the similarly expressed genes in patients with polycystic ovary syndrome and transsexuals

Rong Dong, MD[a], Shang Gao, MB[b], Meng-Jie Shan, MD[c,d,*] [iD]

Abstract

Polycystic ovary syndrome (PCOS) is a common female infertility, which may be caused by excessive androgen, but its mechanism remains unknown. Transsexuals are women who take androgen drugs for a long time, and gradually have male signs. Their ovaries may have received high concentrations of androgen, which leads to the failure of ovarian reproductive function. Therefore, we searched the relevant data of PCOS and transsexuals in gene expression omnibus database, used limma package to identify the most similarly genes, and then analyzed the possible mechanism of PCOS through gene ontology (GO) and kyoto encyclopedia of genes and genomes (KEGG) pathway analysis. Then, the protein-protein interaction network was constructed by searching the String database, and the top 5 hub genes were identified by the cytohubba plug-in of Cytoscape. Finally, ubiquitin conjugating enzyme E2 E1 (UBE2E1), ubiquitin C (UBC), transcription elongation factor B subunit 1 (TCEB1), ubiquitin conjugating enzyme E2 N (UBE2N), and ring finger protein 7 (RNF7) genes were identified as the most similarly expressed genes between PCOS and Transsexuals. They may cause the ubiquitination of androgen receptor and eventually lead to sinus follicular growth arrest. In conclusion, 5 Central genes were identified in PCOS and transsexuals. These genes can be used as targets for early diagnosis or treatment of PCOS.

Abbreviations: GO = gene ontology, KEGG = kyoto encyclopedia of genes and genomes, POCS = polycystic ovary syndrome, RNF7 = ring finger protein 7, SEGs = similar expressed genes, TCEB1 = transcription elongation factor B subunit 1, UBC = ubiquitin C, UBE2E1 = ubiquitin conjugating enzyme E2 E1, UBE2N = ubiquitin conjugating enzyme E2 N.

Keywords: bioinformatics, polycystic ovary syndrome, transsexuals

1. Introduction

Polycystic ovary syndrome (PCOS) is a highly heterogeneous endocrine disorder common in women of childbearing age.[1] PCOS can cause fertility difficulties for women of childbearing age,[2] and it almost affects the physiological state of women throughout their lives.[3] The basic features of PCOS include ovulation dysfunction, hyperandrogen performance, and polycystic changes of the ovary. PCOS can lead to type 2 diabetes, atherosclerosis, and cardiovascular disease.[4] The etiology of PCOS is still unclear. The disease may be associated with genetic factors, chromosomal abnormalities, gene fusion and other factors. It is important to study the molecular mechanism of PCOS.

Transgender refers to a person who has undergone surgery to change his or her natural sex.[5] From the perspective of psychology, "transsexuals" are the denial of their own gender and external genitalia, thus requiring the transformation of their biological gender characteristics, which is called "transsexuals" in psychology. Medically, a transsexual is a person who lives in a changed gender role by undergoing gender reassignment surgery.[6] They believe that they are of the opposite sex inside, claiming to be a member of the opposite sex, requiring medical science to change the body into the sex it identifies.[7] They want people around them to accept them as they want to, these are main characteristics of transgender people. With the progress of the society, more and more transgender people come into people's sight, and people also begin to accept this group who has been drifting in the marginalized society. Long-term hormone

therapy after SEGeneration may also have an impact on the body of transgender people.[8] Antonio's studies have shown that the prevalence of polycystic ovary syndrome (POCS) in transgender people is higher than that in the control group, and the symptoms are very similar.[9] Bioinformatics is an interdisciplinary subject aimed at a comprehensive theory and tool, using a variety of disciplines to clarify and interpret huge biomedical data that contains the biological significance, thus to reveal and understand molecular mechanism of disease occurrence.[10] As the continuous accumulation of biological data resources with the characteristics of big data and the opening of precision medicine strategic plan, the importance of bioinformatics is becoming increasingly obvious.

Therefore, this study intends to use bioinformatics technology to explore the similarly genes between patients with PCOS and transsexuals (female to male), and then to explore the core genes. Thus to investigate the molecular mechanism of POCS.

2. Materials and methods

2.1. Dataset

Gene expression omnibus (http://www.ncbi.nlm.nih.gov/geo) is the largest public database including much biological tissue expression chip and sequence data. The search strategy is (POCS or Polycystic Ovary Syndrome) and ([Female to Male] or Transsexual or treatment with Testosterone). GSE87435 was selected and submitted by Arne IJpma in 2016, including microarray expression data derived from ovaries of women with POCS and Female to Male Transsexual (TSX) individuals after treatment of Testosterone.

2.2. Identification of similarly expressed genes

The datasets were sequenced on the platform of [HG-U133A] Affymetrix Human Genome U133A Array, after downloading the annotation Platform file, the expression matrices extracted from GSE87435 were annotated with R software (version 4.0.0), According to the annotation profiles, probes were annotated, and idle probes without gene symbols were excluded. The average of the probe sets of values were calculated as the expression values for the same gene with multiple probe sets. And then the datasets were normalized by limma package (version 3.44.3),[11] series matrix files were then converted. These similarly expressed genes (SEGs) in POCS and TSX group samples were analyzed by limma package, BH method was used for multiple test correction. |$\log_2$ fold change (FC)| <0.05 and a corrected P value >.8 were used as the cut-off criteria of SEGs analysis. Package ggplot2 (version 3.3.3) and the package pheatmap (version 1.0.12) were used for the volcano plot and heatmaps of SEGs.[12]

2.3. Pathway and functional enrichment analysis

The gene ontology (GO) analysis consists of cellular components, biological processes and molecular functions.[13] The kyoto encyclopedia of genes and genomes (KEGG) database contains various biological pathways.[14] Through GO and KEGG enrichment analysis, it is easy to understand the corresponding pathways of related genes and the molecular mechanism of the disease. In this study, the enrichment analysis of GO and KEGG was mainly conducted with the cluster Profiler package (v3.16.0) in the R software,[15] online comparison analysis of the KEGG database (significant as q-value <0.05 and P < .05).

2.4. Protein-protein interaction network construction and analysis

String database is an online database of protein and protein interactions. We explored the relationship between potential differential genes on STRING (http://www.string-db.org/) version 11.[16] The interaction score > 0.9 is considered to be significant. After deleting disconnected nodes in the network, the relationship was visualized using Cytoscape software. Cytohubba, one plug-in of Cytoscape, can explore the hub genes in the protein-protein interaction network.[17] We chose the MCC, MNC, EPC, and DEGREE algorithms to explore the hub genes.

2.5. Identification of chemicals associated with hub genes

The comparative toxicology genomics database (CTD; http://ctdbase.org/) is a major public resource for literature-based, artificially planned linkages between chemistry, genetic products, phenotypes, diseases, and environmental exposure.[18] We searched the database for these hub genes and looked for chemicals associated with polycystic ovary syndrome.

3. Result

3.1. Identification of similar expressed genes

After downloading the expression matrix and clinical data, samples meeting the inclusion criteria were selected. Finally, the expression data of 6 PCOS ovaries and 12 TSX ovaries were selected. After standardization and normalization of this microarray data, 1609 SEGs between POCS and TSX were extracted from GSE87435, as shown in the volcano plots (Fig. 1A). The heatmap showed that there was no significant difference in the expression of random 50 genes between 2 groups (Fig. 1B). Finally, the 1609 SEGs were used in the following analysis.

3.2. Gene ontology enrichment analysis of similar expressed genes

GO and KEGG analysis were performed to further investigate the biological functions of the 1609 SEGs. The results of GO term enrichment analysis vary with GO classifications, as shown in Figure 2A. In the biological process, SEGs were significantly enriched in the establishment of organelle localization, positive regulation of cellular protein localization, regulation of cell-substrate adhesion, stress granule assembly, mitotic nuclear division, etc. For cellular components, SEGs were enriched in secretory granule lumen, cytoplasmic vesicle lumen, vesicle lumen, spindle, non-motile cilium, photoreceptor outer segment, etc. Aa for molecular functions, no related pathways were enriched. More detailed GO enrichment analysis results are shown in Supplementary Table S1, http://links.lww.com/MD/G377.

3.3. Kyoto encyclopedia of genes and genomes enrichment analysis of similar expressed genes

The SEGs enriched a total of 28 pathways, as shown in the Supplementary Table S2, http://links.lww.com/MD/G378. The KEGG pathway analysis of the top 10 pathways are shown in Figure 2B. The result of this analysis is different from the GO term enrichment analysis, indicating that there are quite complex molecular mechanisms in POCS. These SEGs were mainly enriched in vascular smooth muscle contraction, cGMP-PKG
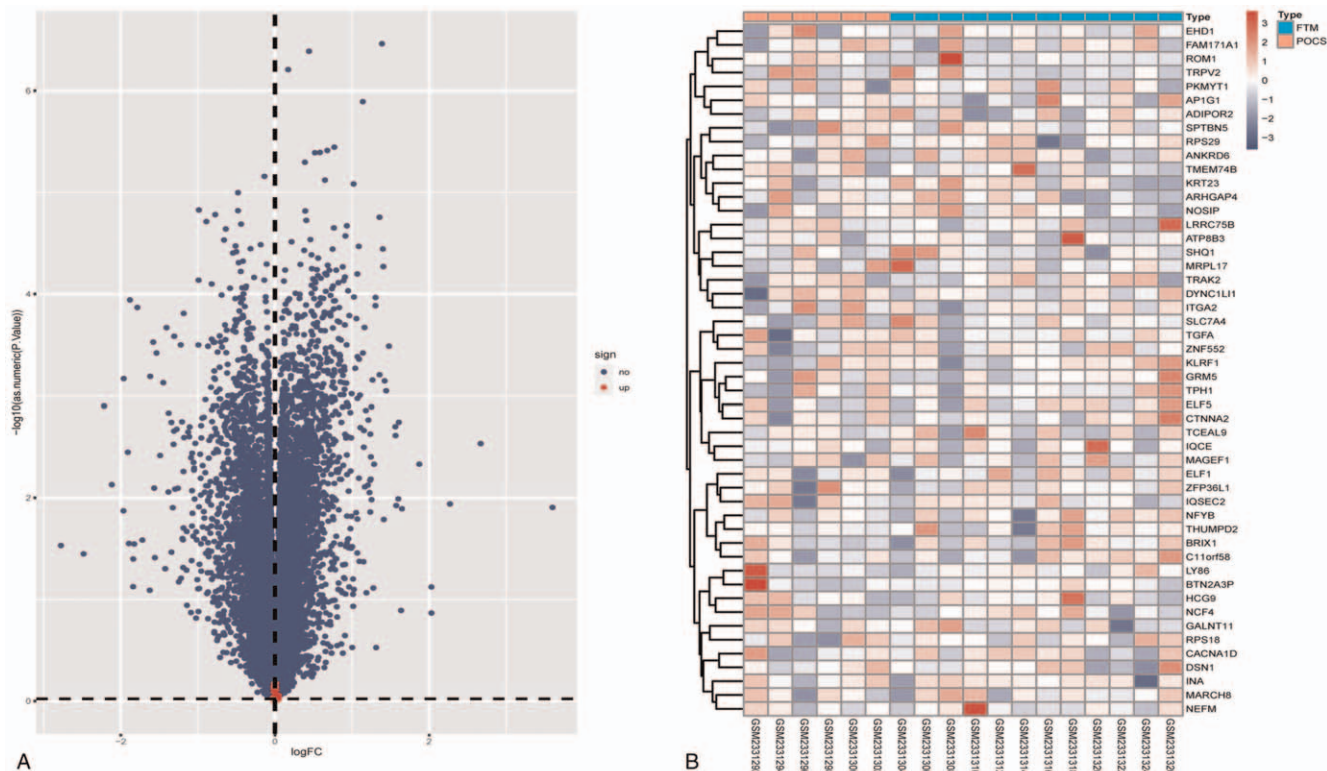
Figure 1. (A). Volcano plots of similarly expressed genes; differentially expressed genes are marked in blue, similarly genes are marked in red; (B). the heatmap of random 50 genes with similarly expressed genes in polycystic ovary syndrome and Transsexuals; Upregulated genes are marked in red, downregulated genes are marked in blue.

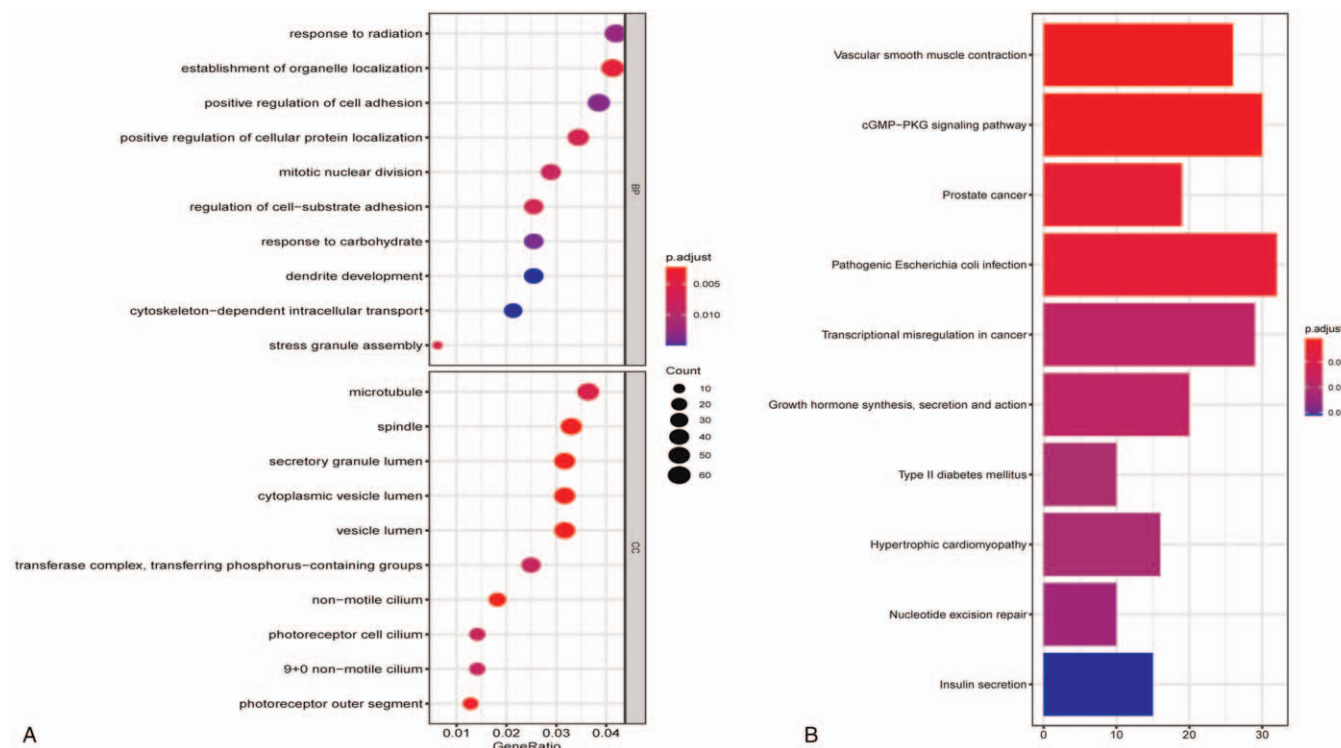**Figure 2.** (A). Enrichment of similarly expressed genes in GO; Different colored circles indicate different adjusted *P* values. The size of the circle indicates the gene count. The y-axis represents the GO term, the x-axis represents the gene proportion; B. Enrichment of similarly expressed genes in KEGG, X-axis represents gene count, Y-axis represents different pathways, and different colors indicate different adjusted P values.
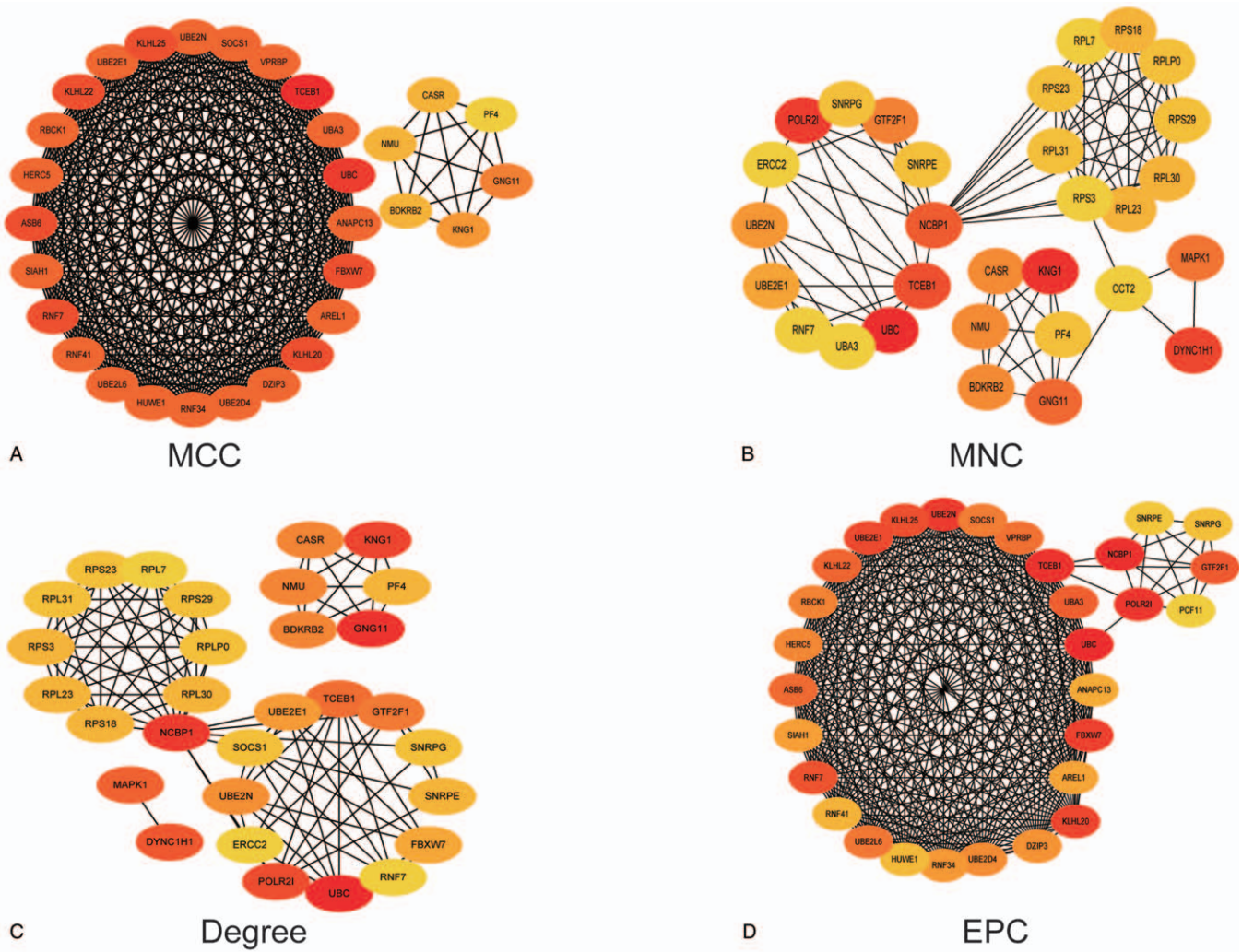
Figure 3. Four different screening methods for the top 30 core genes were selected. (A). MCC methods; (B). MNC methods. C. Degree methods. D. EPC methods.

signaling pathway, Growth hormone synthesis, secretion and action, Type II diabetes mellitus, AMPK signaling pathway, Fatty acid elongation, etc.

3.4. Construction of similar expressed gene protein-protein interaction network

All SEGs were used to construct protein interaction networks. The interaction results were displayedusing Cytoscape software. The MCC, MNC, EPC, and DEGREE methods were used to screen out the hub genes by the CytoHubba plug-in, as shown in Figure 3, they are the top 30 SEGs among different methods. At last, a Venn diagram was plotted (as shown in Fig. 4) to show the intersections among 4 methods, we identified 5 genes that might be related to polycystic ovary syndrome, including ubiquitin conjugating enzyme E2 E1 (UBE2E1), ubiquitin C (UBC), transcription elongation factor B subunit 1 (TCEB1), ubiquitin conjugating enzyme E2 N (UBE2N), and ring finger protein 7 (RNF7).

3.5. Chemicals associated with hub genes

We analyzed each of the hub genes and predicted the chemicals that might affect their expression in POCS. For instance, Ethinyl,
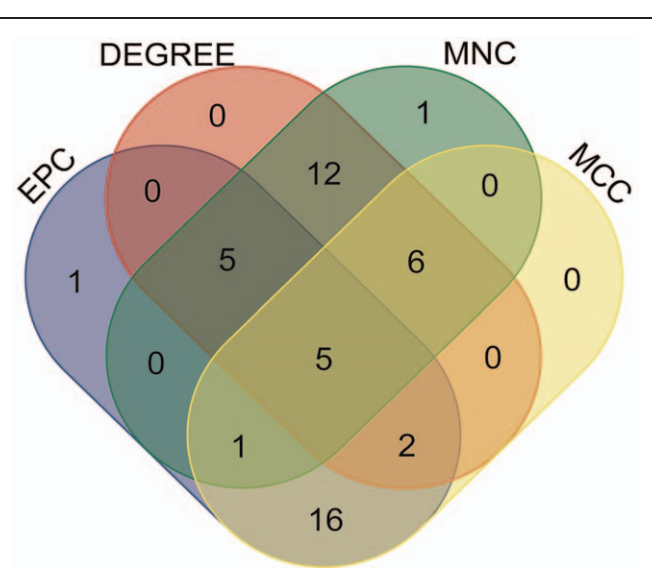

Figure 4. The Venn map of the top hub genes of 4 algorithms (EPC, Degree, MNC, MCC).

Table 1

The chemical associated with polycystic ovary syndrome through the hub genes.

Num	Gene symbol	Gene name	Chemical	Inference score
1	UBE2E1	Ubiquitin Conjugating Enzyme E2 E1	Ethiny, Estradiol, Folic Acid, Valproic Acid	8.62
2	UBC	Ubiquitin C	Estradiol, Ethinyl Estradiol, Flutamide, Folic Acid, Resveratrol, Testosterone, Valproic Acid, Zinc	20.5
3	UBE2N	Ubiquitin Conjugating Enzyme E2 N	Flutamide, Folic Acid, Valproic Acid	6.76
4	RNF7	Ring Finger Protein 7	Estradiol, Ethinyl Estradiol, Flutamide, Folic Acid, Valproic Acid	16.07

Estradiol, Folic Acid, and Valproic Acid may affect the expression of UBE2E1. The rest detailed results are shown in Table 1.

4. Discussion

PCOS is a common gynecological endocrine disorder characterized by hyperandrogen syndrome, insulin resistance, ovulation dysfunction, and polycystic ovary change. Its clinical manifestation is highly heterogeneous. Infertility is one of the main symptoms of PCOS patients in childbearing age.[19] As a common endocrine disease in clinical gynecology, the incidence of PCOS accounts for 5% to 10% of women of childbearing age. Infertility is one of the main symptoms of PCOS patients in childbearing age.[20] Studies have shown that PCOS is the most common cause of ovulation dysfunction and is associated with an increased risk of infertility. PCOS is also a risk factor of cardiovascular disease. Classical risk factors include hypertension, dyslipidemia, diabetes and obesity, etc. Non-classical risk factors include the increase of c-reactive protein, homocysteine and tumor necrosis factor,[21] which can occur at any age in PCOS patients. Patients often have an abnormal distribution of upper body fat, even if they are not accompanied by obesity or have a normal BMI. Visceral fat accumulation and elevated fasting insulin levels are considered as key factors in the development of metabolic diseases in PCOS.

It is very important to explore the molecular mechanism of PCOS in early diagnosis and treatment. The enrichment analysis of KEGG and GO showed that PCOS may be related to vascular smooth muscle contraction, cGMP PKG signaling pathway, growth hormone synthesis, secretion and function, type 2 diabetes, AMPK signaling pathway and fatty acid metabolism.

In this study, bioinformatics analysis was used to conclude that UBE2E1, UBC, TCEB1, UBE2N and RNF7 genes were similarly expressed genes between patients with PCOS and the transsexuals (female to male). UBE2E1 accepts ubiquitin from the E1 complex and catalyzes its covalent attachment to other proteins, and catalyzes the covalent attachment of ISG15 to other proteins,[22] mediating the selective SEGradation of short-lived and abnormal proteins. It also catalyzes "Lys-48"-linked polyubiquitination in vitro. Polge et al found that UBE2E1 is preferably expressed in the cytoplasm of slow-shrinking fibers and protects skeletal muscle from worsening atrophy following dexamethasone treatment, andUBE2E1 may play an important role in skeletal muscle atrophy.[23] They found that the expression of E2E1 was limited to type I and IIA muscle fibers and could not be detected in type IIB muscle fibers. This strongly suggests that target protein of E2E1 is fiber-specific and may be closely related to the contractile and metabolic characteristics of skeletal muscle. UBE2E1 may have protective effect on muscles. UbE2E1 has not been studied in the literature focusing on the PCOS. We have discovered this new molecule which needs to be further explored and may become a new therapeutic target.

The UBC gene plays a key role in maintaining ubiquitin homeostasis. Studies have shown that increasing ubiquitin levels may control UBC gene expression by affecting the splicing of its pre-mRNA, providing a direct feedback strategy for the homeostasis control of ubiquitin library.[24] Lim JJ studied the regulatory mechanism of ubiquitination of androgen receptor signaling in PCOS. The hyperandrogen regulates RNF6 levels and subsequent androgen receptor ubiquitination, leading to antral follicular growth stagnation.[25] This study has established a bridge for UBC, ubiquitination and PCOS. In the future, the pathogenesis of UBC affecting PCOS through ubiquitination could be studied. Agell et al found that TCEB1 promotes invasion of prostate cancer cells.[26] There is no research on TCEB1 in the field of PCOS. UBE2N is a K63 specific ubiquitin binding enzyme associated with many immune diseases and cancers. Dikshit et al demonstrated that UBE2N is highly expressed in melanoma and promotes melanoma growth via MEK/FRA1/SOX10 Signaling.[27] The research of UBE2N is mainly focused on the field of melanoma. There is no research on UBE2N in the field of PCOS. RNF7, also known as RBX2 (RING box protein 2), SAG (Sensitive to Apoptosis Gene), was characterized as the second member of the RBX/ROC RING component of the E3 ubiquitin ligase, SAG promotes ubiquitination and degradation of a number of protein substrates.[28] Some studies have shown that RNF7 may be associated with some cancers,[29] but it has not been reported withPCOS.

To sum up, UBE2E1, UBC, TCEB1, UBE2N and RNF7 are rarely studied in the field of PCOS, which need to be further explored and may be used as new therapeutic targets in the future.

Despite the rigorous bioinformatics analysis in this paper, there are still some deficiencies. In this study, no animal experiments of gene overexpression or knockout were conducted to further verify its function. Therefore, in-depth exploration of core genes should be carried out in the future research.

5. Conclusion

In this study, bioinformatics analysis was used to conclude that UBE2E1, UBC, TCEB1, UBE2N and RNF7 genes were similarly expressed genes between patients with PCOS and the transsexuals (female to male), which may serve as therapeutic targets for PCOS and help us understand its pathogenesis.

Author contributions

Conceptualization: Meng-jie Shan.
Investigation: Rong Dong, Shang Gao.
Software: Rong Dong.

Writing – original draft: Rong Dong.
Writing – review & editing: Meng-jie Shan, Shang Gao.

References

[1] Meier RK. Polycystic ovary syndrome. Nurs Clin North Am 2018; 53:407–20.

[2] Holton S, Hammarberg K, Johnson L. Fertility concerns and related information needs and preferences of women with PCOS. Hum Reprod Open 2018;2018:hoy019.

[3] Jin P, Xie Y. Treatment strategies for women with polycystic ovary syndrome. Gynecol Endocrinol 2018;34:272–7.

[4] Ibáñez L, Oberfield SE, Witchel S, et al. An international consortium update: pathophysiology, diagnosis, and treatment of polycystic ovarian syndrome in adolescence. Horm Res Paediatr 2017;88: 371–95.

[5] Connolly MD, Zervos MJ, Barone CJ, Johnson CC, Joseph CL. The mental health of transgender youth: advances in understanding. J Adolesc Health 2016;59:489–95.

[6] Winter S, Diamond M, Green J, et al. Transgender people: health at the margins of society. Lancet 2016;388:390–400.

[7] Safer JD, Tangpricha V. Care of the transgender patient. Ann Intern Med 2019;171:ITC1–6.

[8] Zurada A, Salandy S, Roberts W, Gielecki J, Schober J, Loukas M. The evolution of transgender surgery. Clin Anat 2018;31:878–86.

[9] Becerra-Fernández A, Pérez-López G, Román MM, et al. Prevalence of hyperandrogenism and polycystic ovary syndrome in female to male transsexuals. Endocrinol Nutr 2014;61:351–8.

[10] Wooller SK, Benstead-Hume G, Chen X, Ali Y, Pearl F. Bioinformatics in translational drug discovery. Biosci Rep 2017;37:BSR20160180. doi: 10.1042/BSR20160180.

[11] Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015;43:e47.

[12] Wickham H, Chang W, RStudio, ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics. Book of Abstracts, 2016.

[13] Denny P, Feuermann M, Hill DP, Lovering RC, Plun-Favreau H, Roncaglia P. Exploring autophagy with gene ontology. Autophagy 2018;14:419–36.

[14] Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res 2017;45(D1):D353–61.

[15] Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 2012; 16:284–7.

[16] Szklarczyk D, Morris JH, Cook H, et al. The STRING database in 2017: quality-controlled protein-protein association networks; 1; made broadly accessible. Nucleic Acids Res 2017;45(D1):D362–8.

[17] Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: identifying hub objects and sub-networks from complex interactome. BMC Syst Biol 2014;8(Suppl 4):S11.

[18] Davis AP, Grondin CJ, Johnson RJ, et al. Comparative toxicogenomics database (CTD): update 2021. Nucleic Acids Res 2021;49(D1):D1138–43.

[19] Nandi A, Chen Z, Patel R, Poretsky L. Polycystic ovary syndrome. Endocrinol Metab Clin North Am 2014;43:123–47.

[20] Moghetti P. Insulin resistance and polycystic ovary syndrome. Curr Pharm Des 2016;22:5526–34.

[21] Barthelmess EK, Naz RK. Polycystic ovary syndrome: current status and future perspective. Front Biosci (Elite Ed) 2014;6:104–19.

[22] Wheaton K, Sarkari F, Stanly Johns B, et al. UbE2E1/UBCH6 is a critical in vivo E2 for the PRC1-catalyzed ubiquitination of H2A at Lys-119. J Biol Chem 2017;292:2893–902.

[23] Cécile P, Julien A, Andrea A, et al. UBE2E1 is preferentially expressed in the cytoplasm of slow-twitch fibers and protects skeletal muscles from exacerbated atrophy upon dexamethasone treatment. Cells 2018;7:214.

[24] Bianchi M, Crinelli R, Giacomini E, et al. A negative feedback mechanism links UBC gene expression to ubiquitin levels by affecting RNA splicing rather than transcription. Sci Rep 2019;9:18556. doi: 10.1038/s41598-019-54973-7.

[25] Lim JJ, Lima P, Salehi R, Lee DR, Tsang BK. Regulation of androgen receptor signaling by ubiquitination during folliculogenesis and its possible dysregulation in polycystic ovarian syndrome. Sci Rep 2017; 7:10272. doi: 10.1038/s41598-017-09880-0.

[26] Agell L, Hernández S, Nonell L, et al. A 12-gene expression signature is associated with aggressive histological in prostate cancer: SEC14L1 and TCEB1 genes are potential markers of progression. Am J Pathol 2012;181:1585–94.

[27] Dikshit A, Jin YJ, Degan S, et al. UBE2N promotes melanoma growth via MEK/FRA1/SOX10 signaling. Cancer Res 2018;78:6462–72.

[28] Sun Y, Li H. Functional characterization of SAG/RBX2/ROC2/RNF7, an antioxidant protein and an E3 ubiquitin ligase. Protein Cell 2013; 4:103–16.

[29] Zuo H, Chen L, Li N, Song Q. Identification of a ubiquitination-related gene risk model for predicting survival in patients with pancreatic cancer. Front Genet 2020;11:1659.