

Identification of polycystic ovary syndrome potential drug targets based on pathobiological similarity in the protein-protein interaction network

Hao Huang^{1,*}, Yuehan He^{1,*}, Wan Li¹, Wenqing Wei¹, Yiran Li¹, Ruiqiang Xie¹, Shanshan Guo¹, Yahui Wang¹, Jing Jiang¹, Binbin Chen¹, Junjie Lv¹, Nana Zhang², Lina Chen¹, Weiming He³

¹College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, Heilongjiang Province, China

²Department of Endocrinology and metabolism Laboratory, The Second Affiliated Hospital of Harbin Medical University, Harbin, Heilongjiang Province, China

³Institute of Opto-electronics, Harbin Institute of Technology, Harbin, Heilongjiang Province, China

*These authors have contributed equally to this work

Correspondence to: Lina Chen, **email:** chenlina@ems.hrbmu.edu.cn
Weiming He, **email:** hewm@hit.edu.cn

Keywords: polycystic ovary syndrome, protein-protein interaction network, pathobiological similarity, module, drug target

Received: January 04, 2016

Accepted: April 28, 2016

Published: May 13, 2016

ABSTRACT

Polycystic ovary syndrome (PCOS) is one of the most common endocrinological disorders in reproductive aged women. PCOS and Type 2 Diabetes (T2D) are closely linked in multiple levels and possess high pathobiological similarity. Here, we put forward a new computational approach based on the pathobiological similarity to identify PCOS potential drug target modules (PPDT-Modules) and PCOS potential drug targets in the protein-protein interaction network (PPIN). From the systems level and biological background, 1 PPDT-Module and 22 PCOS potential drug targets were identified, 21 of which were verified by literatures to be associated with the pathogenesis of PCOS. 42 drugs targeting to 13 PCOS potential drug targets were investigated experimentally or clinically for PCOS. Evaluated by independent datasets, the whole PPDT-Module and 22 PCOS potential drug targets could not only reveal the drug response, but also distinguish the statuses between normal and disease. Our identified PPDT-Module and PCOS potential drug targets would shed light on the treatment of PCOS. And our approach would provide valuable insights to research on the pathogenesis and drug response of other diseases.

INTRODUCTION

Polycystic ovary syndrome (PCOS) is one of the most common endocrinological disorders in reproductive aged women, of which the prevalence is 5–10% in the general population [1–3]. The clinical and biochemical characteristics of PCOS are generally grouped into two main categories: (i) reproductive features that mainly include chronic anovulation, and (ii) hyperandrogenic features that mainly include hyperandrogenaemia.

As important metabolism characteristics, insulin resistance and compensatory hyperinsulinemia are observed in approximately 70% of PCOS women [4], which contribute to the strong association of PCOS with adverse metabolic risk, including dysglycemia, dyslipidaemia and fatty liver, as well as increase the risk for development of type 2 diabetes (T2D) [5]. It is found

that some other features including weight gain, disturbance of hormones and lipid disorders are associated with both PCOS and T2D [6, 7]. In the meantime, PCOS and T2D are closely linked in the epidemiology and etiology as well as in the systems biology [8, 9]. Significant pathobiological similarity between PCOS and T2D was also indicated (−0.2) after determining the pathobiological and clinical relationship between these diseases using a method proposed by Jörg Menche et al. [9].

Metformin is the only known drug for PCOS in Drugbank. The DrugBank database is a unique bioinformatics and cheminformatics resource that combines detailed drug data with comprehensive drug target information, which was widely recognized and used [10]. As an insulin sensitizer, Metformin can cause reductions in body weight, restore ovulation, increase the rate of pregnancy and reduce the number of pregnancy

complications [11]. In the meantime, metformin is also used for treating T2D, which improves glycemic control by decreasing hepatic glucose production and glucose absorption and increasing insulin-mediated glucose uptake [12]. In addition, Thiazolidinediones (TZDs), including pioglitazone, are peroxisome proliferator-activated receptor (PPAR) agonists that induce adipogenesis and have insulin-sensitizing and antidiabetic properties [13]. Recent studies have demonstrated the beneficial metabolic effects of treatment with pioglitazone in PCOS patients [11, 14]. These indicated that the genes and drugs associated with common metabolism characteristics of PCOS and T2D might be the potential drug targets and drugs of PCOS.

The research on complex diseases from the systems level and biological background might provide new insights into drug target identification, disease gene exploration and pathogenesis elucidation [15–17]. For example, disease genes and drug targets of coronary heart disease (CAD) and T2D could be identified effectively by systems biology approaches [18, 19]. Chan KH et al. identified key drivers that included COL1A1, COL3A1, and ELN in the shared pathways for both CAD and T2D based on the protein-protein interaction network (PPIN) and genome-wide association studies [20]. In addition, the highly central importance proteins in core networks or sub-networks are globally or locally important nodes which could act as functional modules or pathways or key role genes to uncover the pathogenesis and therapeutic ways of disease [21]. For example, Li H et al. presented a novel PPI knowledge-based approach that identified functional modules and hub genes for CAD, revealing several novel pathogenic mechanisms [20]. Wan Li et al. developed a method using combined centrality indices on the systems level to identify essential cancer-related motifs in central roles, which provided a clue for the study of signal transduction in biological pathways [22]. Furthermore, Emad Fadhal et al. have indicated that the genes possessing highly topologically centered in the PPIN may be good therapeutic targets [23].

In this study, based on the pathobiological similarity of PCOS and T2D, a new computational approach (Figure 1) was performed to identify the PCOS potential drug target modules (PPDT-Modules) and PCOS potential drug targets from the systems level and biological background in the PPIN. Our study may reveal novel PCOS therapeutic ways and improve the cure rate of PCOS.

RESULTS

PPDT-Modules

The PPDT-Module identification approach we put forward was performed with 35 PCOS disease genes, 83 T2D disease genes and 62 T2D drug targets in the PPIN. 910 P-Modules (13 P-Modules contained one or more

PCOS disease genes) and 923 T-Modules (25 T-Modules contained one or more T2D disease genes) were identified. The Jaccard similarity index was calculated for each P-Module and T-Module pairs. 3 candidate PPDT-Modules were identified ultimately, including PCOS disease genes, T2D disease genes and known T2D drug targets simultaneously (Table 1, Supplementary Table 1).

Central importance of genes in PPDT-Modules

Degree, closeness, betweenness, and page rank of genes in PPDT-Modules were calculated, and the genes in each module were ranked by these topological metrics, respectively. The G-rank of genes in PPDT-Modules was calculated using the geometric mean rank (G-rank of top 22 genes and all genes in PPDT-Module 2 are shown in Table 2 and Supplementary Table 2, respectively).

The PPDT-Modules were the pathobiological similar modules of PCOS and T2D. The higher genes of PPDT-Module were ranked, the more important they might be, and the more kernel roles they would act in the pathogenesis and treatment of PCOS. The genes ESR1, RXRA and NCOA1 were the top three genes not only in the G-rank of genes, but also in four individual topological ranks, which demonstrated the stability of PPDT-Module 2. It was noted that ESR1 (1st) and ESR2 (5th) both ranked highly in the G-rank of genes in PPDT-Module 2. They are the main receptors of estrogen and the metabolism of estrogens was proved to be associated with the pathogenesis of PCOS [24]. The gene RXRA (2nd) is always binding with PPARG and function together, and the gene PPARG is both the PCOS disease gene and T2D disease gene, is also the known T2D drug target, and ranked 12th in PPDT-Module 2. The gene NCOA1 (3rd) is one of the known T2D drug targets. In addition, the gene PPARA, ranked 11th in PPDT-Module 2, is one of the known T2D targets. So genes which possessed high G-rank in PPDT-Module might play key roles in the pathogenesis, diagnosis and treatment of PCOS.

PCOS potential drug targets

From the functional perspective to analyze the association between PPDT-Modules and PCOS, the hyper-geometric test was performed with the genes in PPDT-Modules and the PCOS disease genes in the Gene Ontology (GO) and pathways of Kyoto Encyclopedia of Genes and Genomes (KEGG), and the p value was adjusted with FDR ($P < 0.05$, Figure 2A).

The genes of PPDT-Module 1 and PCOS disease genes, PPDT-Module 3 and PCOS disease genes are not enriched significantly in the functional categories related to PCOS, respectively (Figure 2A). The genes of PPDT-Module 2 and PCOS disease genes are enriched significantly in functional categories and KEGG pathways associated with the pathogenesis of PCOS (Figure 2A,

Supplementary Table 3), such as transcription, steroid hormone receptor signaling pathway (GO biological process, BP), steroid binding, lipid binding and steroid hormone receptor binding (GO molecular function, MF), PPAR signaling pathway, Steroid hormone biosynthesis (KEGG pathway).

“GO:0030518~steroid hormone receptor signaling pathway”, “GO:0003707~steroid hormone receptor activity”, “GO:0005496~steroid binding” and “GO:0005496~steroid binding” are associated with steroid hormones. Disturbed steroid hormone receptor signaling pathway would lead to the hormonal imbalance which is a main underlying problem with PCOS [25]. The steroid hormones mainly contain the androgens and estrogens and so on. “GO:0030521~androgen receptor signaling pathway”

and “GO:0050681~androgen receptor binding” are also associated with androgen. Androgen receptor signaling is an important mediator in the etiology of PCOS traits [26]. And androgens mediate their steroidal actions via the androgen receptor which are crucial in maintaining female fertility by optimizing follicle growth, health, and ovulation [27]. With PCOS, women typically have high levels of androgens, high levels of androgens affect the development and release of eggs during ovulation [27–29]. “GO:0030520~estrogen receptor signaling pathway” and “GO:0030331~estrogen receptor binding” are associated with estrogen. It is noted that estrogens play important roles in the development and functioning of the male and female reproductive systems [30, 31]. The main mediators of estrogen action are 2 specific high-affinity receptors, the ESR1 and ESR2, are

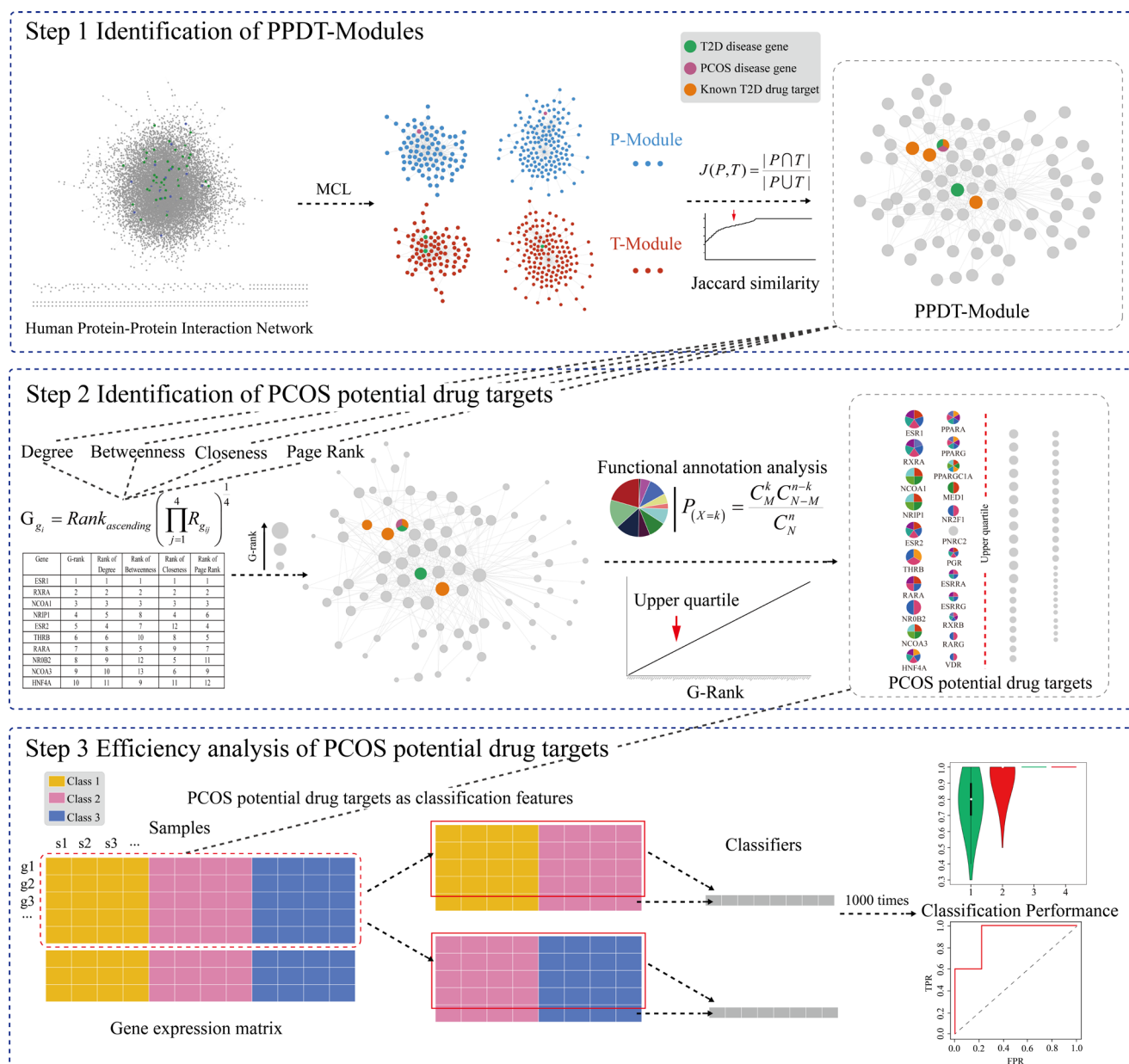


Figure 1: A schematic diagram of PCOS potential drug targets identification and efficiency analysis.

Table 1: Candidate PPDT-Modules

Candidate PPDT-Module	The number of genes in PPDT-Module	PCOS disease Gene	T2D disease Gene	T2D drug Target
Candidate PPDT-Module 1	141	TH	UCP3	PTPN1, ADRA2C
Candidate PPDT-Module 2	85	PPARG	HNF4A, PPARG	NCOA1, PPARA, PPARG, PPARG
Candidate PPDT-Module 3	6	INSR	INSR, ENPP1	INSR

Table 2: G-rank of top 22 genes in PPDT-Module 2

Gene	G-rank	Rank of Degree	Rank of Betweenness	Rank of Closeness	Rank of Page Rank
ESR1	1	1	1	1	1
RXRA	2	2	2	2	2
NCOA1	3	3	3	3	3
NRIP1	4	5	8	4	6
ESR2	5	4	7	12	4
THRB	6	6	10	8	5
RARA	7	8	5	9	7
NR0B2	8	9	12	5	11
NCOA3	9	10	13	6	9
HNF4A	10	11	9	11	12
PPARA	11	12	4	30	10
PPARG	12	7	11	28	8
PPARGC1A	13	13	22	7	13
MED1	14	14	20	10	14
NR2F1	15	28	6	13	24
PNRC2	16	15	19	16	17
PGR	17	19	14	22	15
ESRRA	18	18	21	14	19
ESRRG	19	16	23	39	16
RXRB	20	17	35	27	18
RARG	21	20	26	33	23
VDR	22	21	25	36	21

necessary for the proper function of the hypothalamic–pituitary–ovarian axis and are expressed in the human ovary, where ESR2 is the predominant receptor and its activation enhances folliculogenesis and ovulation [32, 33]. The expression of ESR2 is lower in follicles derived from women with PCOS compared with healthy women, while ESR1 expression is markedly increased in theca cells of polycystic ovaries, causing alteration in the ESR1/ESR2 ratio in PCOS and possibly abnormal follicular development [34, 35]. Otherwise, insulin resistance is associated not only with the pathogenesis of T2D, but also with the pathogenesis of PCOS [36, 37]. The role of ESR1 in the development of insulin resistance has been described in human with a null mutation in the ESR1, leading to unresponsiveness to estrogen, who developed impaired glucose tolerance [34, 38].

“GO:0019216~regulation of lipid metabolic process”, “GO:0004879~ligand-dependent nuclear receptor activity” and “GO:0008289~lipid binding” are associated with lipid metabolism. Abnormal lipid metabolism is one of the main metabolic characteristics of PCOS patients [39]. The parameters of lipid metabolism were significantly different in women with PCOS from in healthy women, higher triglycerides (TG), and very-low-density lipoprotein cholesterol (VLDL) with lower high-density lipoprotein cholesterol (LDL) in PCOS [40, 41]. Altered expression of genes involved in lipid metabolism could contribute to insulin resistance in non-obese PCOS patients [39]. Based on case-controlled studies, women with PCOS appear to have a high prevalence of the metabolic syndrome including obesity and dyslipidemia, and have a higher risk of diabetes [2]. Insulin resistance

appears to have a pivotal role with multiple studies demonstrating significant associations between altered lipid metabolism and insulin level, which may be induced in part by the insulin-medicated stimulation of lipolysis and altered expression of lipoprotein lipase and hepatic lipase [42, 43]. Some studies indicated that disturbance of transcription, disturbance of steroid hormone, lipid abnormalities and so on should lead to promoting the PCOS [44–46].

Meanwhile, “hsa03320:PPAR signaling pathway” plays an important role in the pathogenesis of PCOS and T2D, and the treatment for T2D. Peroxisome proliferator-activated receptors (PPARs) are nuclear hormone receptors activated by fatty acids and their derivatives [13]. PPAR signaling pathway was shown to associate with the pathogenesis of PCOS, as well as with the pathogenesis of T2D [13]. PPAR has three subtypes (PPARA, PPARD, and PPARG), all of them are known T2D drug targets, are showing different expression patterns. PPARA plays a role in the clearance of circulating or cellular lipids via the regulation of gene expression involved in lipid metabolism in liver and skeletal muscle [47]. PPARD is involved in lipid oxidation and cell proliferation [47, 48]. PPARG promotes adipocyte differentiation to enhance blood glucose uptake [49]. Meanwhile, PPARG is both PCOS and T2D disease gene, the drug targeting to PPARG was used to treat with PCOS patient and had good curative effect [13].

These changes of function and disturbance of pathways were associated not only with PCOS, but also with T2D closely. So, the genes of PPDT-Module 2 enriched significantly in these functions, may act as PCOS and T2D disease genes and drug targets (Figure 2B), such as PPARG. Known T2D drug targets NCOA1 and PPARA, T2D disease genes HNF4A, and the PCOS disease genes AR, CYP11A1 were also enriched in same functional categories and pathways of PPARG. Meanwhile, these disease genes and drug targets in PPDT-Module 2 had a high G-Rank, for instance, NCOA1 ranked 3rd, HNF4A ranked 10th, PPARA ranked 11th, PPARG ranked 12th. Hence, 22 genes of PPDT-Module 2 whose G-Rank was the upper quartile as well as enriched significantly in these PCOS associated functional categories and pathways were identified as the PCOS potential drug targets (Figure 3, Table 2).

Efficiency analysis of PCOS potential drug targets

To assess the efficiency of PCOS potential drug targets, SVM was applied to classify samples of different statuses (normal/PCOS, PCOS/after pioglitazone treatment) of GSE8157 with the average expression values of identified PCOS potential drug targets as classification features. 1000 times of five-fold cross-validation were performed (Figure 4). With the average expression values of 22 PCOS potential drug targets as the classification

features, the average of sensitivity values for normal/PCOS, PCOS/after pioglitazone treatment were 0.79 and 1, the average of specificity values were 0.68 and 1, the average of AUCs were 0.72 and 1, the average of accuracy values were 0.73 and 1, the average of MCC scores were 0.49 and 1, respectively (Figure 4, Figure 5A, Supplementary Table 5, and Supplementary Table 6). At the same time, the average expression values of PCOS disease genes, T2D disease genes and known T2D drug targets were used as the classification features to classify the same samples of different statuses, respectively (Figure 4, Supplementary Table 5, and Supplementary Table 6). Most of these classification features could classify samples efficiently.

For the original expression profile, expression consistency check was performed via Wilcoxon signed-rank test to test the sample consistency of the same status (13 normal samples, 10 PCOS samples and 10 after pioglitazone treatment samples, FDR P value < 0.05), respectively. 9 of 13 normal samples and 5 of 10 after pioglitazone treatment samples were consistent. These 5 samples may alter to a common status after drug treatment, suggesting similar drug response and good curative effect of drug treatment, so the paired 5 PCOS/after pioglitazone treatment samples were selected and a new SVM was applied to classify the consistent samples of 3 statuses using the above-mentioned classification features (Figure 4). With identified PCOS potential drug targets as classification features, the average of sensitivity values for normal/PCOS, PCOS/after pioglitazone treatment were 0.95 and 1, the average of specificity values for normal/PCOS, PCOS/after pioglitazone treatment were 0.88 and 1, average of AUC values of normal/PCOS and PCOS/after pioglitazone were 0.91 and 1, the average of accuracy values for normal/PCOS, PCOS/after pioglitazone treatment were 0.92 and 1, the average of MCC scores for normal/PCOS, PCOS/after pioglitazone treatment were 0.84 and 1, respectively (Figure 4, Figure 5A, Supplementary Table 5, and Supplementary Table 6). The classification performance of consistent samples was much better than that of all samples from the original profile with the same classification features, especially for using the average expression values of identified PCOS potential drug targets and known T2D disease genes as classification features (Figure 4, Figure 5, Supplementary Table 5, and Supplementary Table 6), which demonstrated the effectiveness of identified PCOS potential drug targets. PCOS potential drug targets could effectively reflect drug response, classify the samples between normal and disease and between disease and after drug treatment, as well as represent the treatment effect to PCOS accurately. Interestingly, identified PCOS potential drug targets had a better classification performance than known PCOS disease genes, which demonstrated that these potential drug targets may act as the PCOS disease genes (Figure 4, Figure 5, Supplementary Table 5, and

Supplementary Table 6). Classification using identified PPDT-Module 2 as classification features had good performance for normal /PCOS and for PCOS/after pioglitazone treatment in both original and consistent samples (Figure 4B, Figure 5B, Supplementary Table 5, and Supplementary Table 6).

DISCUSSION

PCOS is a disorder of irregular menses, hyperandrogenism and/or polycystic ovary morphology [26]. A large proportion of women with PCOS also exhibit insulin resistance, β cell dysfunction, impaired glucose tolerance and/or T2D [50]. Women with T2D have a higher prevalence of polycystic ovary syndrome compared to the general population [8]. There is also a high rate of T2D in family members of women with PCOS [50]. Some studies have demonstrated that there are common biological background and marker between PCOS and T2D [51]. Based on these, we put forward a new method to identify 3 candidate PPDT-Modules from the systems level in the PPIN based on the pathobiological similarity of PCOS and T2D. Identified PPDT-Modules contained many PCOS and T2D disease genes, and known T2D drug targets, simultaneously. By the functional annotation analysis with the PCOS disease genes, candidate PPDT-Module 2 was identified as the final PPDT-Module. The

functional categories and pathways that genes of PPDT-Module 2 and PCOS disease genes enriched significantly, are associated not only with pathogenesis and treatment of T2D, but also with PCOS. Considering the central importance and functional annotation of known disease genes and targets, 22 PCOS potential drug targets were identified. These PCOS potential drug targets could describe the drug treatment effect well by an independent profile verified, and the 21 genes (95.45%) had been verified by literature to be associated with the pathogenesis and treatment of PCOS (Supplementary Table 4).

Dysfunctions of steroid hormone receptor signaling pathway, lipid metabolic process, and fatty acid oxidation are mainly metabolic disturbances of PCOS [44–46]. Identified PPDT-Module 2 had not only a good consistency function with PCOS but also good classification performance for classifying normal/PCOS samples and PCOS/after pioglitazone treatment samples before and after consistent check. These suggested that PPDT-Module 2 responded to drug treatment efficiently, may be used as an integral PCOS potential therapeutic target. 22 PCOS potential drug targets possessed central importance in PPDT-Module 2 as well as had good consistent function with PCOS. It was worth noting that 22 PCOS potential drug targets we identified were all interconnected in PPDT-Module 2, and had good classification performance, as same as one with PPDT-Module2, could represent the whole PPDT-Module 2 well.

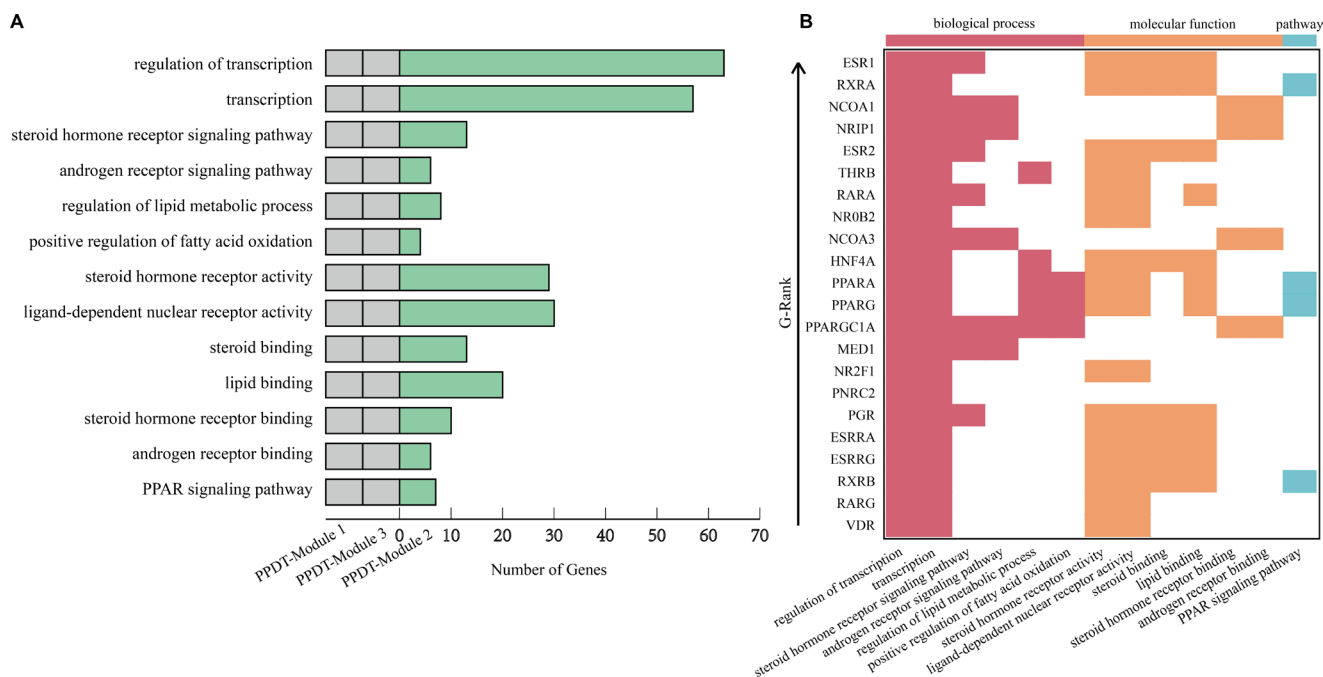


Figure 2: Main functions annotated by PPDT-Modules. (A) Main functions annotated by 3 candidate PPDT-Modules and PCOS disease genes. The first two columns of gray bars represent no gene of candidate PPDT-Module 1 and candidate PPDT-Module 3 enriched significantly in corresponding functional categories, respectively. The third column of green bars represents the number of genes of candidate PPDT-Module 2 enriched in the functional categories. (B) Main functions annotated by genes in PPDT-Module 2. Main functions annotated by genes in PPDT-Module 2. Each row represents a gene in PPDT-Module 2, and each column represents a functional category: biological process, molecular function and pathway. Arrow represents the ascending trend of G-Rank of genes in PPDT-Module 2.

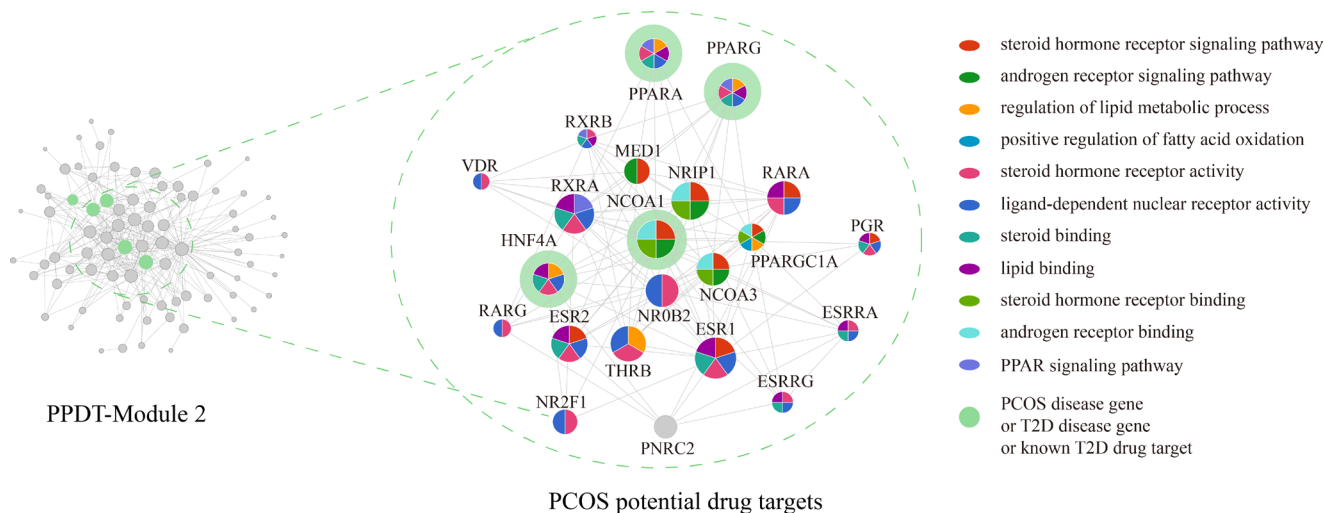


Figure 3: PPDT-module 2 and PCOS potential drug targets. The size of nodes represents their G-Ranks. Different colors represent different functional categories. The green circle represents PCOS disease gene or T2D disease gene or known T2D drug target.

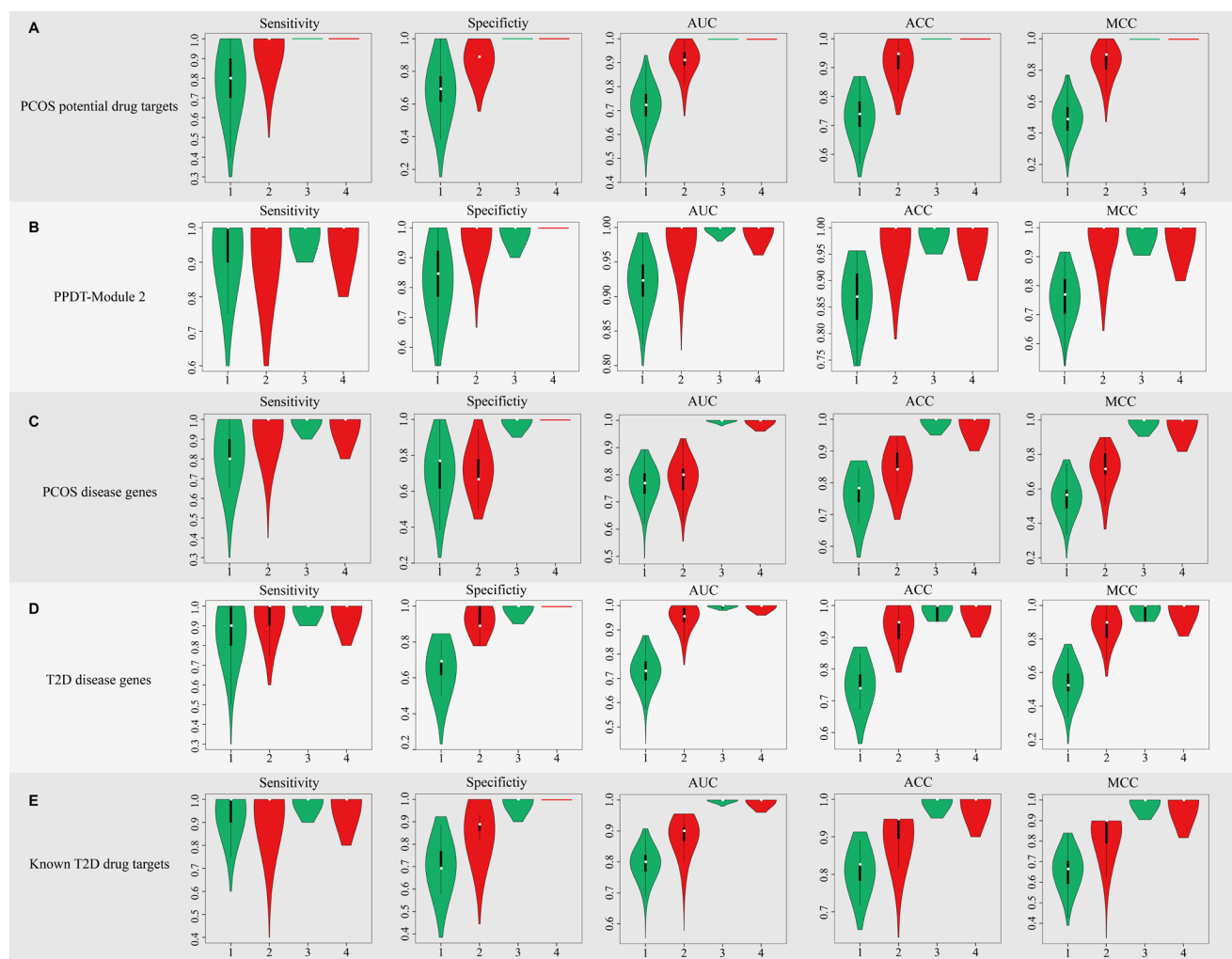


Figure 4: The sensitivity, specificity, AUC, ACC and MCC score distribution of different classification features with 1000 times five-fold cross-validation. In each figure, the first two boxes represent the distribution of classification of normal/PCOS samples, the next two represent the distribution of classification of PCOS/after pioglitazone treatment samples. Green represents the classification of the samples before consistency check, red represents the classification of the samples after consistency check.

Recent studies have demonstrated that the beneficial metabolic effects of treatment with pioglitazone induce adipogenesis and have insulin-sensitizing and antidiabetic properties in PCOS patients [52, 53]. It is known that different patients' response to the drug treatment are different, thus the efficacy of treatment is various. The consistent check could identify status-consistent samples to improve the identification accuracy of disease status and effectiveness evaluation accuracy of drug response. The classification performances of samples with consistent expressions were much better than that of all samples with the average expression values of PCOS potential drug targets, PPDT-Module 2, PCOS disease genes, T2D disease genes and known T2D drug targets as classification features to classify 3 statuses samples of the expression profile GSE8157, respectively (Figure 4). In the meantime, the classification performances of samples with consistent expressions were also much better than that of all samples of original expression profile with same classification features in another independent expression profile GSE6798 related to PCOS (Supplementary Figure S1). Interestingly, the classification performances of PCOS potential drug targets and PPDT-Module 2 as

classification features were better than PCOS disease genes, T2D disease genes and known T2D drug targets in both expression profiles (Figure 5, Supplementary Table 5, Supplementary Table 6 and Supplementary Table 7).

The potential drug targets and disease genes were usually identified based on significantly differential expressed genes [54]. It was noted that between normal samples and PCOS samples no significantly differential expressed genes of expression profiles GSE8157 and GSE6798 could be identified, while between PCOS samples and after pioglitazone treatment samples of GSE8157 only few significantly differential expressed genes could be identified, including no known disease genes, known drug targets or our interested genes. So the potential drug targets and disease genes could not be identified from the point of differential expressed genes for these two expression profiles. Nevertheless, identified PPDT-Module 2 and 22 PCOS potential drug targets could distinguish the statuses between normal and disease, evaluate the effectiveness of drug treatment from the systems level and functional point. Just like our research on the PCOS from the biological background, previous researchers used the profile GSE8157 to identify

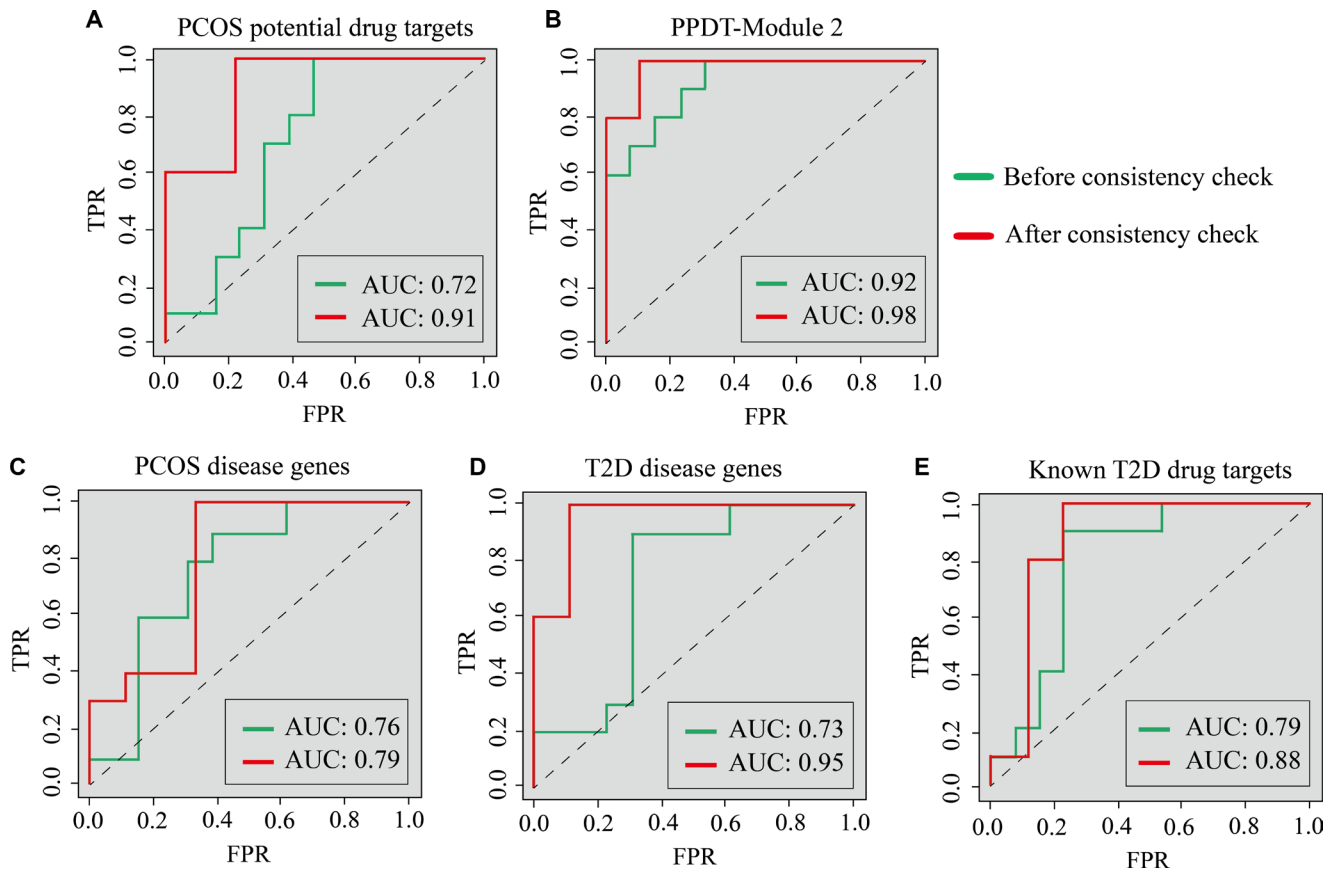


Figure 5: The ROC curves of classification for normal/PCOS with different features. ROC curves of classification for normal/PCOS with (A) PCOS potential drug targets, (B) PPDT-Module 2, (C) PCOS disease genes, (D) T2D disease genes and (E) Known T2D drug targets as classification features, respectively. Green lines represent the ROC curves before consistency check, red lines represent the ROC curves after consistency check.

Table 3: Drugs targeting PCOS potential drug targets

PCOS potential drug targets	Drugs
ESR1	Mestranol ^b , Fulvestrant ^a , Danazol ^a , Etonogestrel ^b , Raloxifene ^b , Diethylstilbestrol ^b , Medroxyprogesterone Acetate ^b , Genistein ^a , Desogestrel ^b , Norgestimate ^b , Naloxone ^b , Levonorgestrel ^a , Ethinyl Estradiol ^b , Melatonin ^a , Clomifene ^b , Progesterone ^b , Tamoxifen ^b , Estradiol ^b , Estrone ^a , Estriol ^a
RXRA	Alitretinoin ^a
NCOA1	Genistein ^a
ESR2	Raloxifene ^b , Diethylstilbestrol ^b , Genistein ^a , Tamoxifen ^b , Estradiol ^b
THRB	Levothyroxine ^a
RARA	Alitretinoin ^a , Isotretinoin ^b
PPARA	Gemfibrozil ^a , Bezafibrate ^b , Indomethacin ^a
PPARG	Repaglinide ^a , Glipizide ^a , Telmisartan ^a , Indomethacin ^a , Pioglitazone ^b , Rosiglitazone ^b , Bezafibrate ^b
PGR	Mifepristone ^a , Danazol ^a , Etonogestrel ^a , Megestrol acetate ^b , Medroxyprogesterone Acetate ^b , Desogestrel ^b , Norgestimate ^b , Levonorgestrel ^a , Drospirenone ^b , Spironolactone ^b , Norethindrone ^b , Progesterone ^b
ESRRG	Diethylstilbestrol ^b
RXRB	Alitretinoin ^a , Tretinoin ^a
RARG	Alitretinoin ^a , Tretinoin ^a
VDR	Alfacalcidol ^a , Calcidiol ^a , Ergocalciferol ^a , Cholecalciferol ^a , Calcitriol ^a

a denotes the drug is under experimental investigation; b denotes the drug has been investigated to be used for treating to PCOS in clinical.

differential pathways instead of differential expressed genes, for analyzing the pathogenesis of PCOS and the way of drug response [52]. It was shown that from the system level and biological background to study the disease statuses and drug response could obtain comprehensive and effective results, especially in the case of no significantly differential expressed genes.

In addition, 21 PCOS potential drug targets were verified by literatures to be associated with the pathogenesis and treatment of PCOS, which could reflect the status of disease and drug response. Drugs targeting these potential drug targets may be potential drugs for PCOS. And 42 drugs targeting to 13 PCOS potential drug targets were verified to be investigated experimentally or clinically for treating PCOS (Table 3). As a drug target, PPARG is widely studied and used for T2D, and is also researched to treat PCOS [55, 56]. There are two classes of drugs targeting to PPARG for T2D. One is insulin sensitization agent such as Pioglitazone, Rosiglitazone; the other is correcting the lipid metabolism disorders agent such as Repaglinide, Telmisartan. These drugs targeting to PPARG were highly possible to act as potential drugs for PCOS. Desogestrel, Ethinyl Estradiol, Clomifene and Tamoxifen are targeting to ESR1 (1st). Desogestrel as synthetic progestational hormone component and Ethinyl Estradiol as estrogenic component in oral contraceptives were used to improve clinical hyperandrogenism in PCOS in the first-line medicine [57, 58]. Clomifene is a triphenyl

ethylene stilbene derivative as an estrogen agonist or antagonist, which is the most commonly used drug for first-line treatment of ovulation induction in women with PCOS [59–61].

Some drugs are targeting several PCOS potential drug targets such as Alitretinoin targeting RXRA, RARA, RXRB and RARG; Tamoxifen targeting ESR1 and ESR2. Retinoid X receptors (RXRs) and retinoic acid receptors (RARs) are nuclear receptors that mediate the biological effects of retinoids by their involvement in retinoic acid-mediated gene activation. These receptors function as transcription factors by binding as homodimers or heterodimers to specific sequences in the promoters of target genes. The proteins encoded by these genes are members of the steroid and thyroid hormone receptor superfamily of transcriptional regulators [62]. Alitretinoin is also known as retinoic acid and derived from maternal vitamin A. Altered retinoic acid synthesis and action could influence the expression of these genes and androgen production in PCOS [63]. Tamoxifen, as one of the selective estrogen receptor modulators (SERM) with tissue-specific activities, is discovered to be a good alternative to clomiphene in women with PCOS and clomiphene-resistant case [64]. These drugs may play key roles in treating PCOS by multiple ways.

As shown in a series of recent publications [16, 65] in developing new findings and approaches, user-friendly and publicly accessible web-servers will significantly

enhance their impacts [20], we shall make efforts in our future work to provide a web-server to displaying findings that can be manipulated by users according to their need.

Our approach was performed based on the existing known diseases genes and drug targets. As the disease genes and drug targets increased, our approach would be more precise. In the meantime, more and more gene expression datasets and novel drugs are available, which will make the result evaluation more objective and precise. The median of the Jaccard similarity indexes was selected as the threshold to identify the pathobiological similar modules in our study, which should be further optimized for the application of other diseases. In addition, the side-effect and toxicology of the drugs identified by identified PCOS potential drug targets were not considered adequately in our study, which will be further improved in our future work.

We proposed a new computational approach to identify the PPDT-Module and PCOS potential drug targets from the systems level and biological background based on the PPIN and pathobiological similarity. 21 genes of 22 identified PCOS potential drug targets were verified to be associated with the pathogenesis of PCOS, and 42 drugs targeting 13 PCOS potential drug targets were investigated experimentally or clinically for PCOS. These PCOS potential drug targets and PPDT-Module could reveal the drug response, distinguish the statuses between normal and disease, and act as PCOS biomarkers. Moreover, further research are expected for the ways of drug response. Our study would shed light on the treatment of PCOS, and provide new insights to research on the pathogenesis and drug response of other diseases that possess high pathobiological similarity.

MATERIALS AND METHODS

Data source

The Protein-Protein Interaction Network was obtained from the Human Protein Reference Database (<http://www.hprd.org/>) [66], which contained 37041 protein-protein interaction pairs between 9518 proteins. Our research was based on two gene expression profiles from the Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo/>) [67] : GSE8157 including 33 samples (13 control samples and 10 PCOS samples before and after pioglitazone treatment, respectively) and GSE6798 including 29 samples (13 control samples and 16 PCOS samples) in GPL570. 35 PCOS disease genes and 83 T2D disease genes were extracted from GAD (<http://geneticassociationdb.nih.gov/>) [68], DO (<http://www.disease-ontology.org/>) [69] and OMIM (<http://www.ncbi.nlm.nih.gov/omim/>) [70]; 62 T2D drug targets were extracted from Drug Bank (<http://www.drugbank.ca/>) [70].

Identification of PPDT-Modules

We put forward a new approach to identify PPDT-modules in the PPIN based on the pathobiological similarity of PCOS and T2D. (i) PCOS disease genes and T2D disease genes were set as seeds to filter the no weight protein-protein interaction network and obtained the disease genes related no weight 3-step-neighbor sub-network, respectively. T2D and PCOS related modules (T-Modules and P-Modules) were mined using the Markov Cluster Algorithm (MCL Algorithm) in the 3-step-neighbor sub-network (no requirement of minimum number of seeds each cluster needs to contain), respectively. The MCL algorithm is a fast and scalable unsupervised cluster algorithm for networks, resulting in a collection of densely connected groups of genes [71]. The Markov clustering parameter was set to the default value 1.8. (ii) The pathobiological similar modules were defined as the intersection of P-Module and T-Module pairs, if their Jaccard similarity indexes were larger than their median, and their intersection contained the PCOS disease genes and T2D disease genes simultaneously. The Jaccard similarity index $J(P, T)$ reflected the consistency of each two modules P-Module and T-Module:

$$J(P, T) = \frac{|P \cap T|}{|P \cup T|} \quad (1)$$

where P is a P-Module and T is a T-Module. The numerator is the number of common genes in P and T. The denominator is the number of all genes in P and T. (iii) Candidate PPDT-Modules were the pathobiological similar modules that included T2D drug targets, and PPDT-Modules were the candidate modules which possessed consistent function with PCOS.

Analysis of the central importance of genes in PPDT-Modules

The genes possessing more neighbors, more paths through them and so on, could play the central role in PPDT-Modules. Hence, the G-rank, G_{gi} , was used to evaluate the central importance of gene g_i in these modules, which integrated 4 topological features, degree, betweenness, closeness and page rank, utilizing the geometric mean rank:

$$G_{gi} = \text{Rank}_{\text{ascending}} \left(\prod_{j=1}^4 R_{g_{ij}} \right)^{\frac{1}{4}} \quad (2)$$

where $R_{g_{ij}}$ is the rank of g_i for the feature j (degree, betweenness, closeness and page rank when $j = 1, 2, 3, 4$, respectively).

The degree of g_i is the number of edges linked to g_i .

The closeness of g_i , C_{gi} , is defined as the inverse of the average geodesic distance (or length of the shortest path) to all other genes:

$$C_{gi} = \left(\frac{1}{(n-1)} \sum_{g_p \neq g_i} d_{g_p g_i} \right)^{-1} \quad (3)$$

where d_{g_p, g_i} denotes the length of shortest path between g_i and g_p , n denotes the number of all genes in the module.

The betweenness of g_i , B_{g_i} , indicates how many of the shortest paths between any two genes in the network include g_i :

$$B_{g_i} = \frac{1}{(n-1)(n-2)} \sum_{g_p \neq g_i, g_q \neq g_i} \frac{\sigma_{g_p, g_i, g_q}}{\sigma_{g_p, g_q}} \quad (4)$$

where σ_{g_p, g_i, g_q} denotes the number of the shortest path between g_p and g_q through g_i , and σ_{g_p, g_q} denotes the number of the shortest paths between g_p and g_q .

The page rank of g_i , PR_{g_i} , should be highly ranked if other highly ranked genes contained edges to it:

$$PR_{g_i} = d \sum_{g_p \neq g_i} \frac{L(g_p)}{PR(g_p)} + \frac{1-d}{n} \quad (5)$$

where $PR(g_p)$ denotes the rank of g_p , $L(g_p)$ denotes the number of edges of g_p , $d \in (0, 1)$ is a fixed parameter.

Identification and analysis of PCOS potential drug targets

The PCOS potential drug targets were defined as the genes with high G-ranks (upper quartile) in PPDT-Modules, as well as correspond to the function of PCOS disease genes (FDR $p < 0.05$). The functional annotation and literature verification were conducted to investigate the correlation between PCOS potential drug targets and PCOS. The functional annotation analysis was performed with all genes in PPDT-Modules and PCOS disease genes using the hyper-geometric test:

$$P_{(X=k)} = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n} \quad (6)$$

where N denotes the number of all human genes, n denotes the number of genes of PPDT-Modules and PCOS disease genes, M denotes the number of genes in function i , k denotes the number of genes of PPDT-Modules and PCOS disease genes in function i . The False Discovery Rate (FDR) method was used to correct p -value, and p -value < 0.05 was set as the criterion for this analysis [72–74].

Efficiency analysis of PCOS potential drug targets

To assess the efficiency of PCOS potential drug targets, the average expression values of them were used as input classification feature values for an independent expression profile of drug treatment (GSE8157, PCOS samples before and after pioglitazone treatment and normal samples) using Support Vector Machine (SVM). In machine learning, SVM is a supervised learning model with associated learning algorithms that analyze data and recognize patterns, which is used for classification and regression analyses [75]. The jackknife test has been increasingly used by investigators to test the power of various predictors, however, to reduce the computational

time, we adopted the five-fold cross-validation as done by many investigators using SVM as the prediction engine [16, 65]. A receiver operating characteristic (ROC) curve was plotted and the area under the curve (AUC), the sensitivity, specificity, accuracy and MCC score were calculated. The five-fold cross-validation was applied 1000 times, and the 1000 ROC curves and AUC scores were obtained to evaluate the robustness of the PCOS potential drug targets. The higher the AUC score was, the better the classification performance was.

ACKNOWLEDGMENTS AND FUNDING

This study was supported by the National Natural Science Foundation of China (Grant No. 31301040 and 61272388); the Health Department Funds of Heilongjiang Province China (Grant No. 2012-810); the Science & Technology Research Project of the Heilongjiang Ministry of Education (Grant No. 12541476); and the Master Innovation Funds of Heilongjiang Province (Grant No. YJSCX2015-40HYD).

CONFLICTS OF INTEREST

The authors declare no potential conflicts of interest.

REFERENCES

1. Allahbadia GN, Merchant R. Polycystic ovary syndrome and impact on health. MEFS. 2011; 16:19–37.
2. Baldani DP, Skrgatic L, Ougouag R. Polycystic Ovary Syndrome: Important Underrecognised Cardiometabolic Risk Factor in Reproductive-Age Women. Int J Endocrinol. 2015; 2015.
3. Thathapudi S, Kodati V, Erukkambattu J, Katragadda A, Addepally U, Hasan Q. Anthropometric and biochemical characteristics of polycystic ovarian syndrome in South Indian women using AES-2006 criteria. Int J Endocrinol Metab. 2014; 12.
4. Reaven GM. Insulin resistance and compensatory hyperinsulinemia: role in hypertension, dyslipidemia, and coronary heart disease. American heart journal. 1991; 121:1283–1288.
5. Barber TM, Dimitriadis GK, Andreou A, Franks S. Polycystic ovary syndrome: insight into pathogenesis and a common association with insulin resistance. Clin Med (Lond). 2015; 15:s72–s76.
6. Corbett SJ, McMichael AJ, Prentice AM. Type 2 diabetes, cardiovascular disease, and the evolutionary paradox of the polycystic ovary syndrome: a fertility first hypothesis. Am J Hum Biol. 2009; 21:587–598.
7. Conn JJ, Jacobs HS, Conway GS. The prevalence of polycystic ovaries in women with type 2 diabetes mellitus. Clin Endocrinol (Oxf). 2000; 52:81–86.

8. Barber TM, Franks S. The link between polycystic ovary syndrome and both Type 1 and Type 2 diabetes mellitus: what do we know today? *Women's Health*. 2012; 8:147–154.
9. Menche J, Sharma A, Kitsak M, Ghiassian SD, Vidal M, Loscalzo J, Barabási A-L. Uncovering disease-disease relationships through the incomplete interactome. *Science*. 2015; 347:1257601.
10. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, Chang Z, Woolsey J. DrugBank: a comprehensive resource for *in silico* drug discovery and exploration. *Nucleic Acids Res*. 2006; 34:D668–D672.
11. Brettenthaler N, De Geyter C, Huber PR, Keller U. Effect of the insulin sensitizer pioglitazone on insulin resistance, hyperandrogenism, and ovulatory dysfunction in women with polycystic ovary syndrome. *J Clin Endocrinol Metab*. 2004; 89:3835–3840.
12. Marchetti P, Del Guerra S, Marselli L, Lupi R, Masini M, Pollera M, Bugliani M, Boggi U, Vistoli F, Mosca F. Pancreatic islets from type 2 diabetic patients have functional defects and increased apoptosis that are ameliorated by metformin. *J Clin Endocrinol Metab*. 2004; 89:5535–5541.
13. Saltiel AR, Olefsky JM. Thiazolidinediones in the treatment of insulin resistance and type II diabetes. *Diabetes*. 1996; 45:1661–1669.
14. Glintborg D, Hermann AP, Andersen M, Hagen C, Beck-Nielsen H, Veldhuis JD, Henriksen JE. Effect of pioglitazone on glucose metabolism and luteinizing hormone secretion in women with polycystic ovary syndrome. *Fertility and sterility*. 2006; 86:385–397.
15. Kitano H. A robustness-based approach to systems-oriented drug design. *Nat Rev Drug Discov*. 2007; 5:202–210.
16. Fu L, Zhang S, Zhang L, Tong X, Zhang J, Zhang Y, Ouyang L, Liu B, Huang J. Systems biology network-based discovery of a small molecule activator BL-AD008 targeting AMPK/ZIPK and inducing apoptosis in cervical cancer. *Oncotarget*. 2015; 6:8071–8088. doi: 10.18632/oncotarget.3513.
17. Levina E, Ji H, Chen M, Baig M, Oliver D, Ohouo P, Lim C-u, Schools G, Carmack S, Ding Y. Identification of novel genes that regulate androgen receptor signaling and growth of androgen-deprived prostate cancer cells. *Oncotarget*. 2015; 6:13088. doi: 10.18632/oncotarget.3743.
18. Watson AD. Thematic review series: systems biology approaches to metabolic and cardiovascular disorders. Lipidomics: a global approach to lipid analysis in biological systems. *J Lipid Res*. 2006; 47:2101–2111.
19. Calimlioglu B, Karagoz K, Sevimoglu T, Kilic E, Gov E, Arga KY. Tissue-specific molecular biomarker signatures of type 2 diabetes: an integrative analysis of transcriptomics and protein-protein interaction data. *OMICS*. 2015; 19:563–573.
20. Chan KHK, Huang Y-T, Meng Q, Wu C, Reiner A, Sobel EM, Tinker L, Lusic AJ, Yang X, Liu S. Shared molecular pathways and gene networks for cardiovascular disease and type 2 diabetes mellitus in women across diverse ethnicities. *Circ Cardiovasc Genet*. 2014; 7:911–919.
21. Safari-Alighiarloo N, Taghizadeh M, Rezaei-Tavirani M, Goliaei B, Peyvandi AA. Protein-protein interaction networks (PPI) and complex diseases. *Gastroenterol Hepatol Bed Bench*. 2014; 7:17.
22. Li W, Chen L, Li X, Jia X, Feng C, Zhang L, He W, Lv J, He Y, Li W. Cancer-related marketing centrality motifs acting as pivot units in the human signaling network and mediating cross-talk between biological pathways. *Mol Biosyst*. 2013; 9:3026–3035.
23. Fadhal E, Gamielien J, Mwambene EC. Protein interaction networks as metric spaces: a novel perspective on distribution of hubs. *BMC Syst Biol*. 2014; 8:6.
24. Bacallao K, Leon L, Gabler F, Soto E, Romero C, Valladares L, Vega M. *In situ* estrogen metabolism in proliferative endometria from untreated women with polycystic ovarian syndrome with and without endometrial hyperplasia. *J Steroid Biochem Mol Biol*. 2008; 110:163–169.
25. Rojas J, Chávez M, Olivar L, Rojas M, Morillo J, Mejías J, Calvo M, Bermúdez V. Polycystic ovary syndrome, insulin resistance, and obesity: navigating the pathophysiological labyrinth. *Int J Reprod Med*. 2014; 2014.
26. Goodarzi MO, Dumesic DA, Chazenbalk G, Azziz R. Polycystic ovary syndrome: etiology, pathogenesis and diagnosis. *Nat Rev Endocrinol*. 2011; 7:219–231.
27. Walters K, Allan C, Jimenez M, Lim P, Davey R, Zajac J, Illingworth P, Handelsman D. Female mice haploinsufficient for an inactivated androgen receptor (AR) exhibit age-dependent defects that resemble the AR null phenotype of dysfunctional late follicle development, ovulation, and fertility. *Endocrinology*. 2007; 148:3674–3684.
28. Sir-Petermann T, Maliqueo M, Angel B, Lara H, Perez-Bravo F, Recabarren S. Maternal serum androgens in pregnant women with polycystic ovarian syndrome: possible implications in prenatal androgenization. *Human Reproduction*. 2002; 17:2573–2579.
29. Barber TM, Wass JA, McCarthy MI, Franks S. Metabolic characteristics of women with polycystic ovaries and oligo-amenorrhoea but normal androgen levels: implications for the management of polycystic ovary syndrome. *Clinical endocrinology*. 2007; 66:513–517.
30. Hyder SM, Stancel GM. Regulation of angiogenic growth factors in the female reproductive tract by estrogens and progestins. *Mol Endocrinol*. 1999; 13:806–811.
31. Gibson DA, Saunders PT. Estrogen dependent signaling in reproductive tissues—a role for estrogen receptors and estrogen related receptors. *Mol Cell Endocrinol*. 2012; 348:361–372.
32. Klinge CM. Estrogen receptor interaction with estrogen response elements. *Nucleic Acids Res*. 2001; 29:2905–2919.

33. Couse JF, Yates MM, Walker VR, Korach KS. Characterization of the hypothalamic-pituitary-gonadal axis in estrogen receptor (ER) null mice reveals hypergonadism and endocrine sex reversal in females lacking ER α but not ER β . *Molecular Endocrinology*. 2003; 17:1039–1053.
34. Nectaria X, Leandros L, Ioannis G, Agathocles T. The importance of ER α and ER β gene polymorphisms in PCOS. *Gynecol Endocrinol*. 2012; 28:505–508.
35. Jakimiuk AJ, Weitsman SR, Yen H-W, Bogusiewicz M, Magoffin DA. Estrogen receptor α and β expression in theca and granulosa cells from women with polycystic ovary syndrome. *J Clin Endocrinol Metab*. 2002; 87:5532–5538.
36. Weyer C, Funahashi T, Tanaka S, Hotta K, Matsuzawa Y, Pratley RE, Tataranni PA. Hypoadiponectinemia in obesity and type 2 diabetes: close association with insulin resistance and hyperinsulinemia. *J Clin Endocrinol Metab*. 2001; 86:1930–1935.
37. Dunaif A. Insulin Resistance and the Polycystic Ovary Syndrome: Mechanism and Implications for Pathogenesis 1. *Endocr Rev*. 1997; 18:774–800.
38. Gallagher CJ, Langefeld CD, Gordon CJ, Campbell JK, Mychalecky JC, Bryer-Ash M, Rich SS, Bowden DW, Sale MM. Association of the Estrogen Receptor- α Gene With the Metabolic Syndrome and Its Component Traits in African-American Families The Insulin Resistance Atherosclerosis Family Study. *Diabetes*. 2007; 56:2135–2141.
39. Chazenbalk G, Chen Y-H, Heneidi S, Lee J-M, Pall M, Chen Y-DI, Azziz R. Abnormal expression of genes involved in inflammation, lipid metabolism, and Wnt signaling in the adipose tissue of polycystic ovary syndrome. *J Clin Endocrinol Metab*. 2012; 97:E765–E770.
40. Macut D, Damjanović S, Panidis D, Spanos N, Glišić B, Petakov M, Rousso D, Kourtis A, Bjekić J, Milić N. Oxidised low-density lipoprotein concentration—early marker of an altered lipid metabolism in young women with PCOS. *Eur J Endocrinol*. 2006; 155:131–136.
41. Swetha R, Ravi B, Nalini K, Abdelazim IA, Elsayah WF, Pillai BP, Prasanna BP, Kumar H, Jayakumar RV, Alur VC. Serum lipoprotein (a) and lipid profile in polycystic ovarian syndrome. *JCSR*. 2015:2.
42. Guilherme A, Virbasius JV, Puri V, Czech MP. Adipocyte dysfunctions linking obesity to insulin resistance and type 2 diabetes. *Nat Rev Mol Cell Biol*. 2008; 9:367–377.
43. Flock MR, Green MH, Kris-Etherton PM. Effects of adiposity on plasma lipid response to reductions in dietary saturated fatty acids and cholesterol. *Methods Mol Biol*. 2011; 2:261–274.
44. Wood JR, Ho CK, Nelson-Degrave VL, McAllister JM, Strauss JF. The molecular signature of polycystic ovary syndrome (PCOS) theca cells defined by gene expression profiling. *J Reprod Immunol*. 2004; 63:51–60.
45. Buyalos R, Geffner M, Watanabe R, Bergman R, Gornbein J, Judd H. The influence of luteinizing hormone and insulin on sex steroids and sex hormone-binding globulin in the polycystic ovarian syndrome. *Fertility and sterility*. 1993; 60:626–633.
46. Diamanti-Kandarakis E, Papavassiliou AG, Kandarakis SA, Chrousos GP. Pathophysiology and types of dyslipidemia in PCOS. *Trends Endocrinol Metab*. 2007; 18:280–285.
47. Berger J, Moller DE. The mechanisms of action of PPARs. *Annu Rev Med*. 2002; 53:409–435.
48. Luquet S, Gaudel C, Holst D, Lopez-Soriano J, Jehl-Pietri C, Fredenrich A, Grimaldi PA. Roles of PPAR delta in lipid absorption and metabolism: a new target for the treatment of type 2 diabetes. *Biochim Biophys Acta*. 2005; 1740:313–317.
49. Yamamoto J, Yamane T, Oishi Y, Shimizu M, Tadaishi M, Kobayashi-Hattori K. Chrysanthemum Promotes Adipocyte Differentiation, Adiponectin Secretion and Glucose Uptake. *Am J Chin Med*. 2015; 43:255–267.
50. Saxena R, Welt C. Polycystic ovary syndrome is not associated with genetic variants that mark risk of type 2 diabetes. *Acta Diabetol*. 2013; 50:451–457.
51. Conway G, Dewailly D, Diamanti-Kandarakis E, Escobar-Morreale HF, Franks S, Gambineri A, Kelestimir F, Macut D, Micic D, Pasquali R. The polycystic ovary syndrome: a position statement from the European Society of Endocrinology. *Eur J Endocrinol*. 2014; 171:P1–P29.
52. Skov V, Glintborg D, Knudsen S, Tan Q, Jensen T, Kruse TA, Beck-Nielsen H, Hojlund K. Pioglitazone enhances mitochondrial biogenesis and ribosomal protein biosynthesis in skeletal muscle in polycystic ovary syndrome. *PLoS One*. 2008; 3:e2466.
53. Dunaif A. Drug insight: insulin-sensitizing drugs in the treatment of polycystic ovary syndrome—a reappraisal. *Nat Rev Endocrinol*. 2008; 4:272–283.
54. Walker MG. Pharmaceutical target identification by gene expression analysis. *Mini Rev Med Chem*. 2001; 1:197–205.
55. Knebel B, Janssen O, Hahn S, Jacob S, Gleich J, Kotzka J, Muller-Wieland D. Increased low grade inflammatory serum markers in patients with Polycystic ovary syndrome (PCOS) and their relationship to PPAR γ gene variants. *Exp Clin Endocrinol Diabetes*. 2008; 116:481–486.
56. Hara M, Alcoser SY, Qaadir A, Beiswenger KK, Cox NJ, Ehrmann DA. Insulin resistance is attenuated in women with polycystic ovary syndrome with the Pro12Ala polymorphism in the PPAR γ gene. *J Clin Endocrinol Metab*. 2002; 87:772–775.
57. Cagnacci A, Paoletti A, Renzi A, Orru M, Pilloni M, Melis G, Volpe A. Glucose metabolism and insulin resistance in women with polycystic ovary syndrome during therapy with oral contraceptives containing cyproterone acetate or desogestrel. *J Clin Endocrinol Metab*. 2003; 88:3621–3625.
58. Bahceci M, Bilge M, Tuzcu A, Tuzcu S, Bahceci S. Serum prostate specific antigen levels in women with polycystic ovary syndrome and the effect of flutamide+ desogestrel/ethinyl estradiol combination. *J Endocrinol Invest*. 2004; 27:353–356.

59. Badawy A, Aal IA, Abulatta M. Clomiphene citrate or letrozole for ovulation induction in women with polycystic ovarian syndrome: a prospective randomized trial. *Fertility and sterility*. 2009; 92:849–852.
60. Legro RS, Barnhart HX, Schlaff WD, Carr BR, Diamond MP, Carson SA, Steinkampf MP, Coutifaris C, McGovern PG, Cataldo NA. Clomiphene, metformin, or both for infertility in the polycystic ovary syndrome. *N Engl J Med*. 2007; 356:551–566.
61. Palomba S, Orio Jr F, Falbo A, Manguso F, Russo T, Cascella T, Tolino A, Carmina E, Colao A, Zullo F. Prospective parallel randomized, double-blind, double-dummy controlled clinical trial comparing clomiphene citrate and metformin as the first-line treatment for ovulation induction in nonobese anovulatory women with polycystic ovary syndrome. *J Clin Endocrinol Metab*. 2005; 90:4068–4074.
62. Beato M, Herrlich P, Schütz G. Steroid hormone receptors: many actors in search of a plot. *Cell*. 1995; 83:851–857.
63. Wood JR, Nelson VL, Ho C, Jansen E, Wang CY, Urbanek M, McAllister JM, Mosselman S, Strauss JF. The molecular phenotype of polycystic ovary syndrome (PCOS) theca cells and new candidate PCOS genes defined by microarray analysis. *J Biol Chem*. 2003; 278:26380–26390.
64. Sowedan I, Salama K, El Shaer O. Tamoxifen versus letrozole in combination with clomiphene citrate to stimulate ovulation in clomiphene-resistant PCOS women. *EBWHJ*. 2015; 5:207–212.
65. Chen J, Wang X, Liu B. iMiRNA-SSF: Improving the Identification of MicroRNA Precursors by Combining Negative Sets with Different Distributions. *Sci Rep*. 2016; 6.
66. Prasad TK, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, Telikicherla D, Raju R, Shafreen B, Venugopal A. Human protein reference database—2009 update. *Nucleic Acids Res*. 2009; 37:D767–D772.
67. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res*. 2002; 30:207–210.
68. Becker KG, Barnes KC, Bright TJ, Wang SA. The genetic association database. *Nat Genet*. 2004; 36:431–432.
69. Schriml LM, Arze C, Nadendla S, Chang Y-WW, Mazaitis M, Felix V, Feng G, Kibbe WA. Disease Ontology: a backbone for disease semantic integration. *Nucleic Acids Res*. 2012; 40:D940–D946.
70. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res*. 2005; 33:D514–D517.
71. Van Dongen S. A cluster algorithm for graphs. *Report-Information systems. CWI*. 2000; 1–40.
72. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol*. 1995; 57:289–300.
73. Willis S, Villalobos VM, Gevaert O, Abramovitz M, Williams C, Sikic BI, Leyland-Jones B. Single Gene Prognostic Biomarkers in Ovarian Cancer: A Meta-Analysis. *PloS one*. 2016; 11:e0149183.
74. Luijk R, Goeman JJ, Slagboom EP, Heijmans BT, van Zwet EW. An alternative approach to multiple testing for methylation QTL mapping reduces the proportion of falsely identified CpGs. *Bioinformatics*. 2015; 31:340–5. doi: 10.1093/bioinformatics/btu654.
75. Witten IH, Frank E. *Data Mining: Practical machine learning tools and techniques*. 2005.