Heliyon 8 (2023) e12480

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Systems biology and *in silico*-based analysis of PCOS revealed the risk of metabolic disorders

Md. Arju Hossain [a], Sheikh Abdullah Al Ashik [a], Moshiur Rahman Mahin [a], Md. Al Amin [a], Md Habibur Rahman [c], Md. Arif Khan [a,b,1], Abdullah Al Emran [a,1,*]

[a] Department of Biotechnology and Genetic Engineering, Mawlana Bhashani Science and Technology University, Tangail, 1092, Bangladesh
[b] Department of Biotechnology and Genetic Engineering, University of Development Alternative, 4/4B, Block A, Lalmatia, Dhaka, 1209, Bangladesh
[c] Department of Computer Science and Engineering, Islamic University, Kushtia, 7003, Bangladesh

HIGHLIGHTS

- The aim of study lies in exploring metabolic risk factors that drive the progression of PCOS based on computational approaches.
- Using global transcriptomic data to identify their genetic profile, pathways, and PPI network analysis for therapeutic intervention.
- Hub protein IGF2R (PDB ID: 2V5O) was employed as a molecular target with certain phytoestrogenic compounds to manage PCOS.
- Integrating molecular docking and the ADMET analysis employed in computational drug discovery as the most inexpensive approach.
- Finally, we have used gold benchmark databases like OMIM and dbGAP to validate the DEGs and molecular pathways.

ARTICLE INFO

Keywords:
Polycystic ovary syndrome
Type 2 diabetes
Cardiovascular disease
Obesity
Metabolic comorbidities
Phytochemicals
ADMET

ABSTRACT

Background: Polycystic ovarian syndrome (PCOS) is a common condition of hyperandrogenism, chronic ovulation, and polycystic ovaries in females during the reproduction and maturation of the ovum. Although PCOS has been associated with metabolic disorders, including type 2 diabetes (T2D), obesity (OBE), and cardiovascular disease (CVD), Causal connection and molecular features are still unknown.

Purpose: Therefore, we investigated the shared common differentially expressed genes (DEGs), pathways, and networks of associated proteins in PCOS and metabolic diseases with therapeutic intervention.

Methods: We have used a bioinformatics pipeline to analyze transcriptome data for the polycystic ovarian syndrome (PCOS), type 2 diabetes (T2D), obesity (OBE), and cardiovascular diseases (CVD) in female patients. Then we employed gene-disease association network, gene ontology (GO) and signaling pathway analysis, selection of hub genes from protein-protein interaction (PPI) network, molecular docking, and gold benchmarking approach to screen potential hub proteins.

Result: We discovered 2225 DEGs in PCOS patients relative to healthy controls and 34, 91, and 205 significant DEGs with T2D, Obesity, and CVD, respectively. Gene Ontology analysis revealed several significant shared and metabolic pathways from signaling pathway analysis. Furthermore, we identified ten potential hub proteins from PPI analysis that may serve as a therapeutic intervention in the future. Finally, we targeted one significant hub protein, IGF2R (PDB ID: 2V5O), out of ten hub proteins based on the Maximal clique centrality (MCC) algorithm and literature review for molecular docking study. Enzastaurin (−12.5), Kaempferol (−9.1), Quercetin (−9.0), and Coumestrol (−8.9) kcal/mol showed higher binding affinity in the molecular docking approach than 19 drug compounds. We have also found that the selected four compounds displayed favorable ADMET properties compared to the native ligand.

Conclusion: Our *in-silico* research findings identified a shared molecular etiology between PCOS and metabolic diseases that may suggest new therapeutic targets and warrants future experimental validation of the key targets.

* Corresponding author.
E-mail address: emrangeb@gmail.com (A.A. Emran).
[1] Joint senior author.

https://doi.org/10.1016/j.heliyon.2022.e12480
Received 4 February 2022; Received in revised form 18 July 2022; Accepted 13 December 2022

1. Introduction

PCOS (polycystic ovarian syndrome) is a complex endocrine disease that negatively influences fertility, metabolism, social characteristics, and the psychological properties of pregnancies, teenage and post-menopausal women [1, 2]. The condition is characterized by hyper-androgenism, ovulatory dysfunction, and morphological functions of the polycystic ovary [3]. The global prevalence rate of polycystic ovary syndrome (PCOS) is 5.6–21.3% and 6% for women of reproductive age and teenagers, respectively [4].

Though the pathogenesis of PCOS is still not well described, it is recognized as a complex condition with distinct metabolic, genetic, environmental, and endocrine aberrations [5]. Molecular mechanisms of PCOS have been increasingly investigated. Wang et al. indicated that the hypoxia-inducible factor (HIF)-1α-mediated endothelin (ET)-2 signaling was closely associated with the progression of PCOS, which was confirmed through the PCOS mouse model [6]. Dysregulated genes were considered to play roles in PCOS associated metabolic dysfunction and follicular growth arrest in patients either with or without insulin résistance [7]. Furthermore, there is evidence that PCOS progression may be influenced by Notch signals and MAPK pathways [8].

Consequently, recent important metabolic facets of PCOS were recognized in connection with longer-term safety sequelae [9]. The metabolic syndrome is a group of diseases characterized by abdominal obesity, insulin resistance, impaired glucose metabolism, hypertension, and dyslipidemia [10]. In particular, obese or overweight women with PCOS are more likely to have IR, which is accompanied by compensatory hyperinsulinemia that contributes to the development of several phenotypic aspects of PCOS [11]. When combined with beta-cell dysfunction, obesity raises the likelihood of developing additional metabolic abnormalities such as type 2 diabetes (T2D), and cardiovascular diseases [12]. The combination of overweight or obesity, T2D worsens not only metabolic, but also reproductive problems linked with this endocrinopathy [13]. Teede et al. conducted a case-controlled study and mentioned that a high incidence of metabolic disorders, including obesity, type 2 diabetes, and cardiovascular disease, were strongly associated with PCOS [14]. For cancer perspectives, only endometrial cancer is 2.7 times more likely in women with polycystic ovarian syndrome (PCOS) [15]. There is conflicting research about PCOS and the risk of ovarian cancer [16]. On the other hand, based on current research, there is no link between PCOS and breast cancer [17]. Above mentioned etiology, among all the metabolic disorders, we have chosen only T2D, obesity and CVD for sharing greatest common female phenotypes to PCOS patients.

Type 2 diabetes mellitus (T2DM), containing 90–95% of patients, is the most common type of diabetes mellitus [18], and at the end of 2030, the figure will increase to 439 million [19]. Several risk factors, including environmental and genetic factors, were co-related in T2DM. Lack of first-stage insulin secretion, irregular basal insulin release, and stimulating glucagon release plays an essential role in developing T2DM [20, 21]. Several studies showed that PCOS and T2D are closely inter-connected [22]. Weerakiet S et al. and Talbott EO et al. conducted separate studies that demonstrated compared with the general population, PCOS is increased the risk of T2DM or impaired glucose tolerance (IGT) than among the general population [23, 24]. At the molecular level, two genes *TCF7L2* and *KCNJ11*, are closely associated with T2DM and PCOS formation [25, 26, 27].

The related morbidity and mortality of obesity are recurrent public-health issues. According to WHO, one of the fifth adults was estimated to be obese at the end of 2025 [28]. Although little information is available, PCOS contributed to the development of obesity and abnormal reproductive growth; J. Vrbikova and V. Hainer (2009) concluded that Obesity and PCOS are linked, and about 30–70% of women who carried PCOS seemed to be obese [29]. Another study conducted by Escobar-morreale F (2005) stated that 17 out of the 36 morbid obese premenopausal women affected PCOS, increasing the risk of obesity [30].

The leading cause of premature death globally, cardiovascular disease (CVD), is a coronary and circulatory disorder, with 23.6 million people dying every year by 2030 [31]. Cardiometabolic, including hypertension, obesity, metabolism disorder, and type II diabetes, is closely associated with PCOS, further increasing CVD risk [32, 33] Woodward A et al. showed PCOS women to carry visceral fat, which could increase chances for a high concentration of oxidized LDL [34] that consequently caused the CVD development [35].

Since there is no actual remedy for PCOS, therapy, and treatments personalized to clinical manifestations are suggested. Liu et al. conducted two separate studies, found some optimistic agents for PCOS intervention, and identified the transcription factor-microRNA [36, 37]. A recent study of Haoran et al., through the pathway and network-based analyses with RT-PCR validation and found *ARHGAP4, ARHGAP9, RHOG*, and *LYN* dysregulated genes that may be involved in the pathogenesis of PCOS [38]. However, several studies were performed between the PCOS and metabolic syndrome, but a network and in silico based approach is not available.

The molecular processes linked to phenotypes of interest can be facilitated using network-based data to illustrate a concise view of the genes or proteins in the system [39].

Investigating PCOS molecular mechanisms that increase the risk of metabolic disorders can help understand the pathogenesis of these diseases with experimental medication design through the microarray and bioinformatics data analysis. As a result, the present study identifies potential pathways based on differentially expressed genes, and protein-protein interactions in PCOS and metabolic disorders. Additionally, an *in silico* approach was performed to develop a putative inhibitor against the specific target. Finally, the research will usher a new avenue on PCOS and metabolic disease molecular pathways with tailored treatment.

2. Materials and methods

2.1. Collection of data

We obtained and examined the pertinent microarray datasets from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus in order to evaluate the effects of PCOS and its genetic correlation with other common disorders. In our study, four different microarray datasets from the Gene Expression Omnibus (GEO) database of the NCBI were examined, including those for PCOS, T2D, obesity (OB), and CVD. These datasets had the accession numbers GSE5850, GSE16415, GSE48521, and GSE24519, respectively.

During the data collecting and analysis process, we exclusively included female samples and excluded male samples considering sex-biased effect of PCOS. We selected the above mentioned datasets compared to the other available datasets due to the following compelling reasons. At the beginning, we filtered the datasets to choose those that

Table 1. Details information of the selected microarray data.

Disease Name	GEO Number	Tissue Name	Control Samples	Case Samples	Significant Genes	Up-regulated Genes	Down-regulated Genes
PCOS	GSE5850	MII arrested oocyte	6	6	2230	949	1344
T2D	GSE16415	Omentum tissue	5	5	908	377	534
Obesity	GSE48521	Amniotic fluid	8	8	1628	803	858
CVD	GSE24519	Platelets	4	10	3220	1760	1461

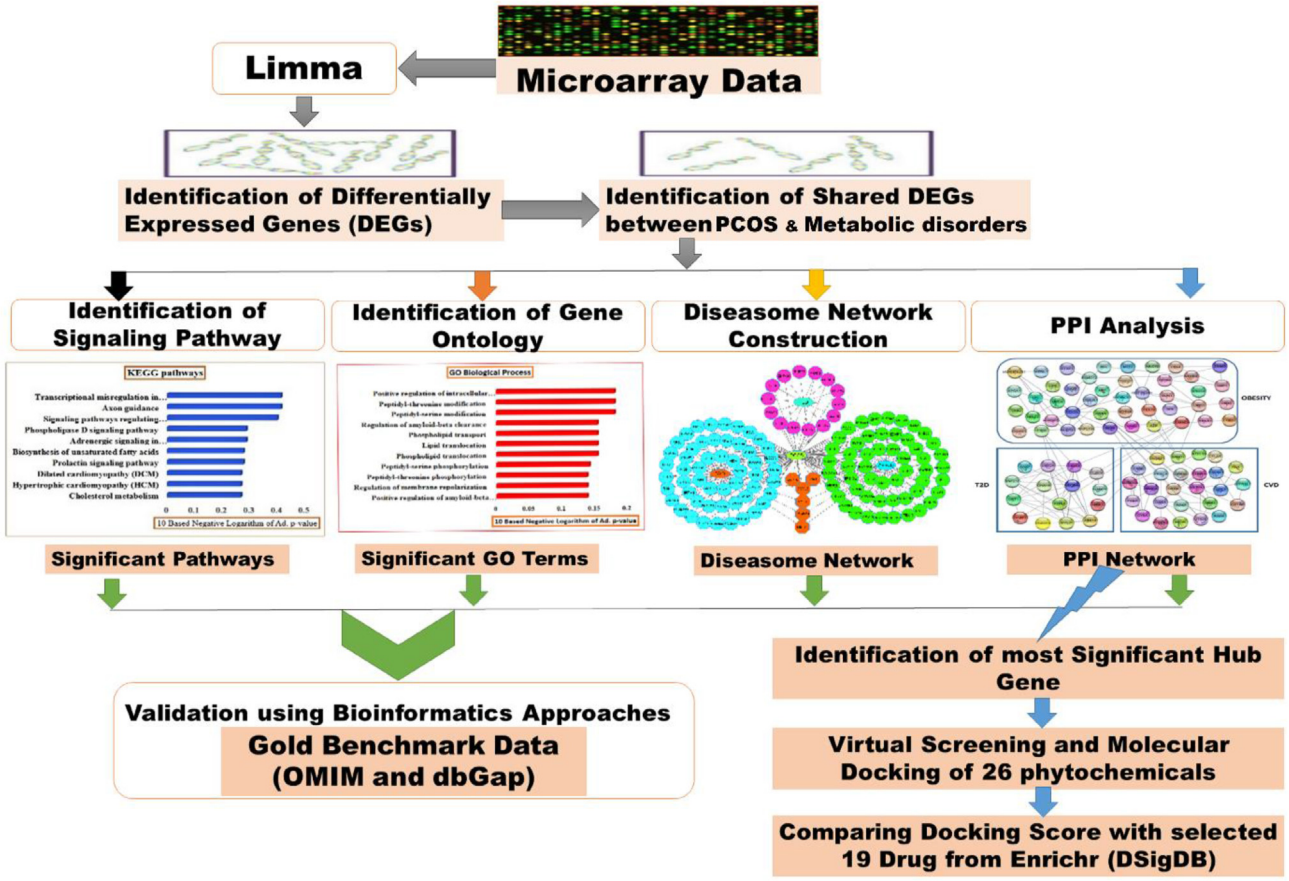


Figure 1. Flow diagram of the analytical approach used in this study. The Gene Expression (GEO) library contains datasets on gene expression. These datasets were evaluated to discover shared differentially expressed genes (DEGs) among the several datasets. Enrichment analysis was used to identify significantly enriched pathways and Gene Ontology (GO) concepts. The network of protein-protein interactions was examined to find hub proteins. Virtual screening and molecular docking were performed to validate hub genes.

had the least bias and noise. Selection criteria, such as sample sizes, gender, sources of tissue or cells, etc., did not line up with the information we had gathered. In our disease datasets, various cells are connected. However, we only consider complete samples that are either case or control and are not related to the regulation. A number of datasets have been labeled with details pertaining to a particular condition, with a focus on biological interactions in particular, although the results do not pertain to the diagnosis and are therefore incorrect. Non-human datasets were removed and only human data was used. For this reason, only one gene expression dataset has been analyzed for the different disease conditions. Details information are presented in Table 1.

2.2. Analysis of differential expression and shared DEGs

We used freely available microarray datasets in this investigation. Two criteria, such as absolute log2 fold change (logFC $>$ 1.5) and absolute P-value (P-value $<$ 0.05) were used in the NCBI GEO2R web tool to identify differentially expressed genes (DEGs) in each dataset.

All four datasets are taken from different cell types/tissues because our study did not depend on tissue-specific gene expression patterns. We standardized the gene expression data including disease state and control state by applying the quantile normalization and Z-score transformation approaches in order to generate consistent expression data from diverse platforms and prevent the issues of experimental systems.

Based on previously published studies [40, 41, 42], we have performed several statistical analyses to remove bias and noise from datasets including two-way analysis of variance (ANOVA) test to determine the statistical significance between groups. P-values were adjusted by the well-established Benjamini-Hochberg method and Bonferroni correction method as indicated. Based on varying the False discovery rate (FDR) threshold and standard statistical criteria; we considered absolute P-value $<$ 0.05 and logfold-change $>$ 1.5 for up-regulated genes, while absolute P-value $<$ 0.05 and logfold-change $<$ -1.5 for finding down-regulated genes. Then we have provided the rationale and justification for the selection of common DEGs by hypergeometric tests. We performed hypergeometric tests for the DEGs to establish their role as predictive diagnostic biomarkers for PCOS and metabolic diseases. Finally, we applied a logarithmic treatment to all of the datasets in order to approximate normality and reduce the impact of outliers.

We have employed to find common DEGs among several diseases by pair-pair common in response to several publications [41, 43] using Venny v2.1 web tool [44]. Venn diagrams are frequently used in biology to illustrate differences across gene lists derived from various differential studies. Thus, they enable comparisons across various experimental setups or methodologies. Multiple programs have been established to create Venn diagrams for use in diverse scientific disciplines [45] including bioinformatics [46], system biology and network modeling [46]. The gene–disease network (GDN) was created and visualized with Cytoscape v3.8.2 [47].

2.3. Pathway and functional correlation analysis

Molecular pathways and gene ontology are assessed via the EnrichR web-based gene-set analysis tool for both PCOS and metabolic disease genes [48]. For gene ontologies, the research included biological process (2018), cellular component (2018), and molecular function (2018); for signaling pathway databases, we investigated Reactome (2016), KEGG

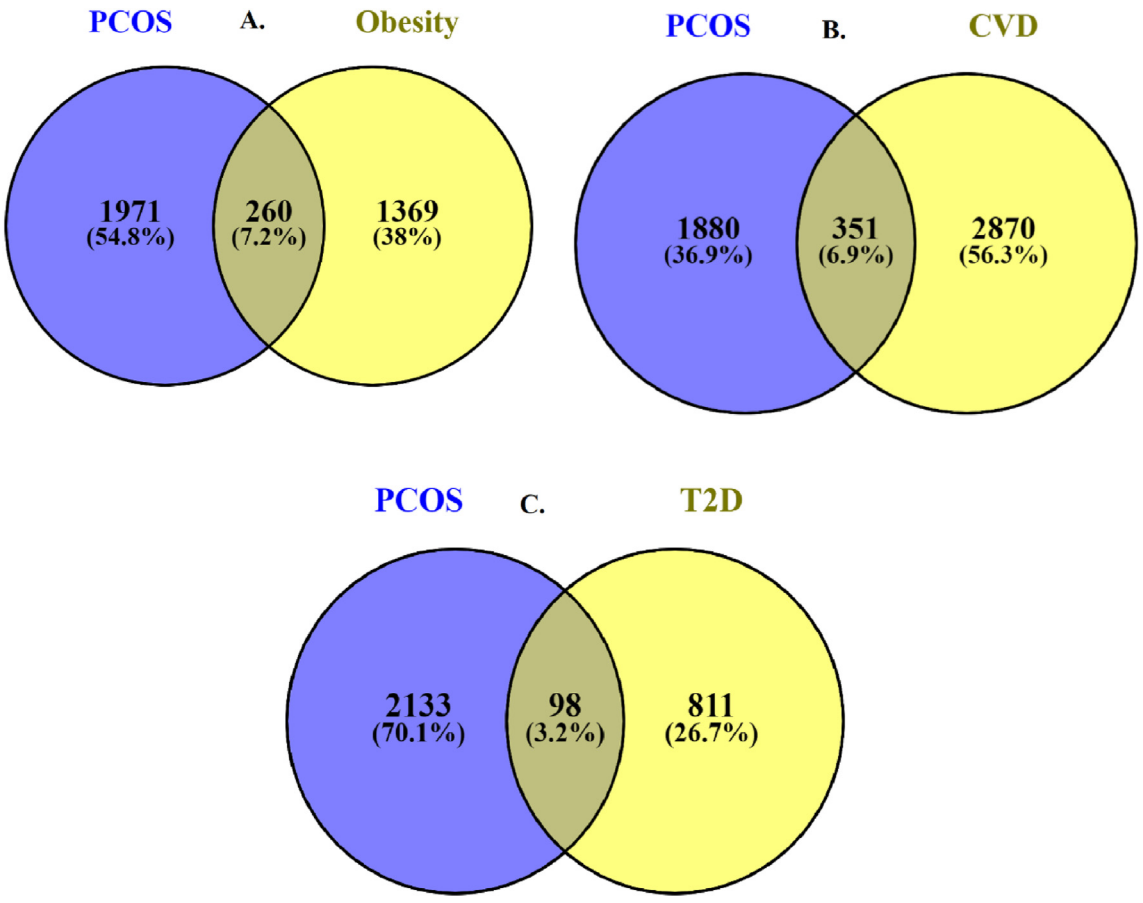


Figure 2. Venn diagram of PCOS and metabolic comorbidities where PCOS shared common dysregulated gene between **A.** Obesity, **B.** Cardiovascular disease (CVD) and **C.** Type 2 diabetes (T2D).

pathways (2019), and Wiki Pathways (2019) databases. The modified P-value was used to filter out the significant routes, and the threshold score was set at 0.05. Using Microsoft Excel 2013, the enrichment plots were displayed as a bar diagram.

2.4. Establishment protein-protein interaction network

For the study of the protein-protein interaction (PPI), we utilized the STRING database [49] for common shared DEGs between PCOS and other metabolic comorbidities via a visualization software Cytoscape (v3.8.2) [47]. The physical connections between proteins in a cell that are created by biochemical reactions or intricate biological processes are known as protein-protein interaction networks. Understanding the physiology of cells in both healthy and diseased states is crucial. To classify strongly interconnected proteins (i.e., hub proteins), a topology study was conducted applying the Cyto-Hubba plugin, and the maximal clique centrality (MCC) methods were implicated out of 12 methods [50]. The new suggested technique, MCC leads the other eleven methods in terms of precision in predicting important proteins from the PPI network. Besides, the most successful approach of locating hub nodes in a co-expression network was revealed to be the Maximal Clique Centrality (MCC) algorithm [50, 51]. Several previously published studies only used MCC methods to identify hub genes out of 12 different approaches [52, 53, 54].

2.5. In silico approach of drug targets

For further studies on molecular docking, the most significant protein in the network of protein-protein interactions was selected [55]. It is an area of dynamic study with many applications in structural medicine architecture, lead optimization, and biochemical routes analysis [56].

2.5.1. Protein and ligand preparation

The protein, Insulin-like Growth Factor Receptor type-2 (IGF2R), was prepared by retrieving the three-dimension crystal structure (PDB: 2V5O) [57] from RCSB PDB [58]. The 3D structure of a target protein was developed by extracting water via Discovery Studio and Pymol software package [59, 60] and minimized total energy with GROMOS96 43B1 force field through SWISS PDB Viewer [61]. We have used an online tool called Computed Atlas of Surface Topography of Proteins (CASTp) to locate, define, and measure these geometrical and topological characteristics of protein structures [62]. We have developed a list of 25 phytochemicals for type 2 diabetes, PCOS, obesity and CVD as possible treatment option from searching of literature review. On the other hand, 19 drug compounds also identified by searching the target protein against several compounds in databases and after screening of hundreds of compounds based on p value sorting, we were able to identify most likely potential compounds. Both ligand molecules were extracted in SDF format from the PubChem database [63]. Pymol Software was used to convert all compounds from SDF to PDB format. Then we have used PyRx integrated mmff94 for optimizing and preparing ligands (Merck molecular force field) [64, 65].

2.5.2. Molecular docking

PyRx has performed molecular docking to explain the ligand and the drug compounds [64]. During ducking, the ligand and target protein were treated as flexible and rigid, respectively. The grid settings tool has been developed using the AutoDock grid box containing the grid box's center points were (X = 67.2844, Y = 25.1326, Z = 27.2148) and size were (X = 133.8326, Y = 126.6655, Z = 73.9831). For ADMET assessment, ligands with the lowest RMSD values and the highest negative docking scores were considered. Finally, the docked pose for molecular interactions between ligands and receptors was examined using Discovery studio and Pymol software's.

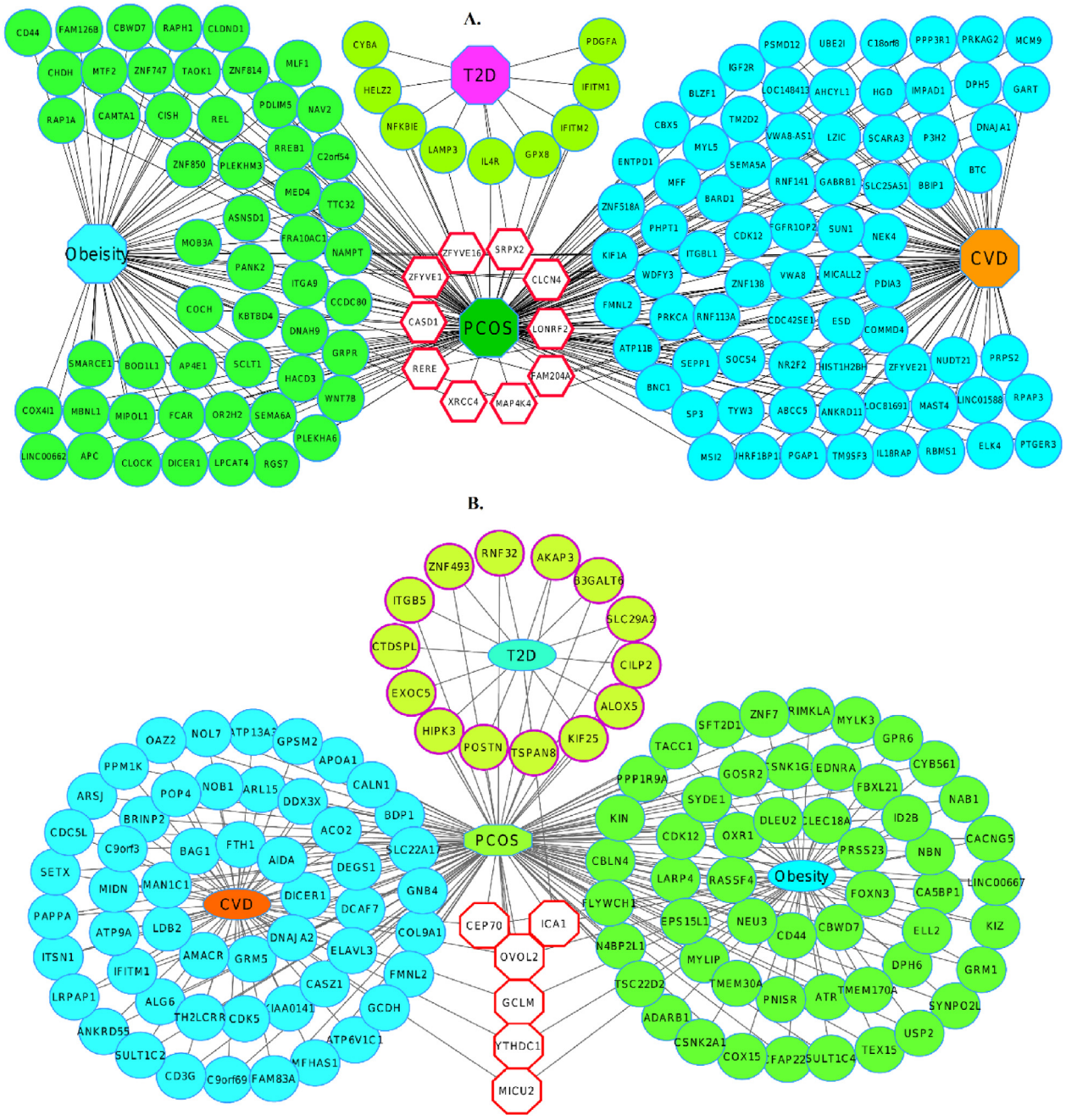


Figure 3. Gene- Disease network of common DEGs having. (A) Up-regulated genes between polycystic ovarian syndrome (PCOS) and type 2 diabetes (T2D), obesity (OBS), and cardiovascular disease (CVD). Octagonal-shaped demonstrate the four Diseases. Circular-shaped indicate the DEGs between PCOS and CVD, T2D and obesity. Hexagonal shaped and red-border color indicate the common DEGs among PCOS, CVD, T2D and Obesity. (B) Down-regulated genes: Ellipse and hexagonal-shaped indicate the metabolic comorbidities. Circular-shaped indicate the DEGs between PCOS and CVD, T2D and obesity. Octagonal-shaped and red-border color indicate the common DEGs among PCOS, CVD, T2D and Obesity.

2.5.3. Analysis of ADMET properties

Higher binding affinity compounds have been chosen for the ADMET analysis. The ADME properties (Absorption, Distribution, Metabolism, and Elimination) of potent drug candidates were measured through SwissADME [66]. To investigate toxicology properties, pkCSM toxicity prediction tools have been used [67].

2.6. Overview of analytical approach

After obtaining 5 microarray datasets from NCBI, we evaluated the datasets using GEO2R's web tools. We performed cross-comparative analysis to identify DEGs from each dataset, then built a gene-disease network, pathway analysis, PPI analysis, identified hub genes, and used an *in-silico* technique. In these experiments, the multi-step quantitative approach as an integrated bioinformatics and system biology pipeline has been developed and applied, as shown in Figure 1.

3. Results

3.1. Differentially expressed genes (DEGs) analysis and distribution of DEGs

In our analysis, we gathered MII-arrested oocytes from ovulating women with normal conditions (control) and PCOS (disease) who were

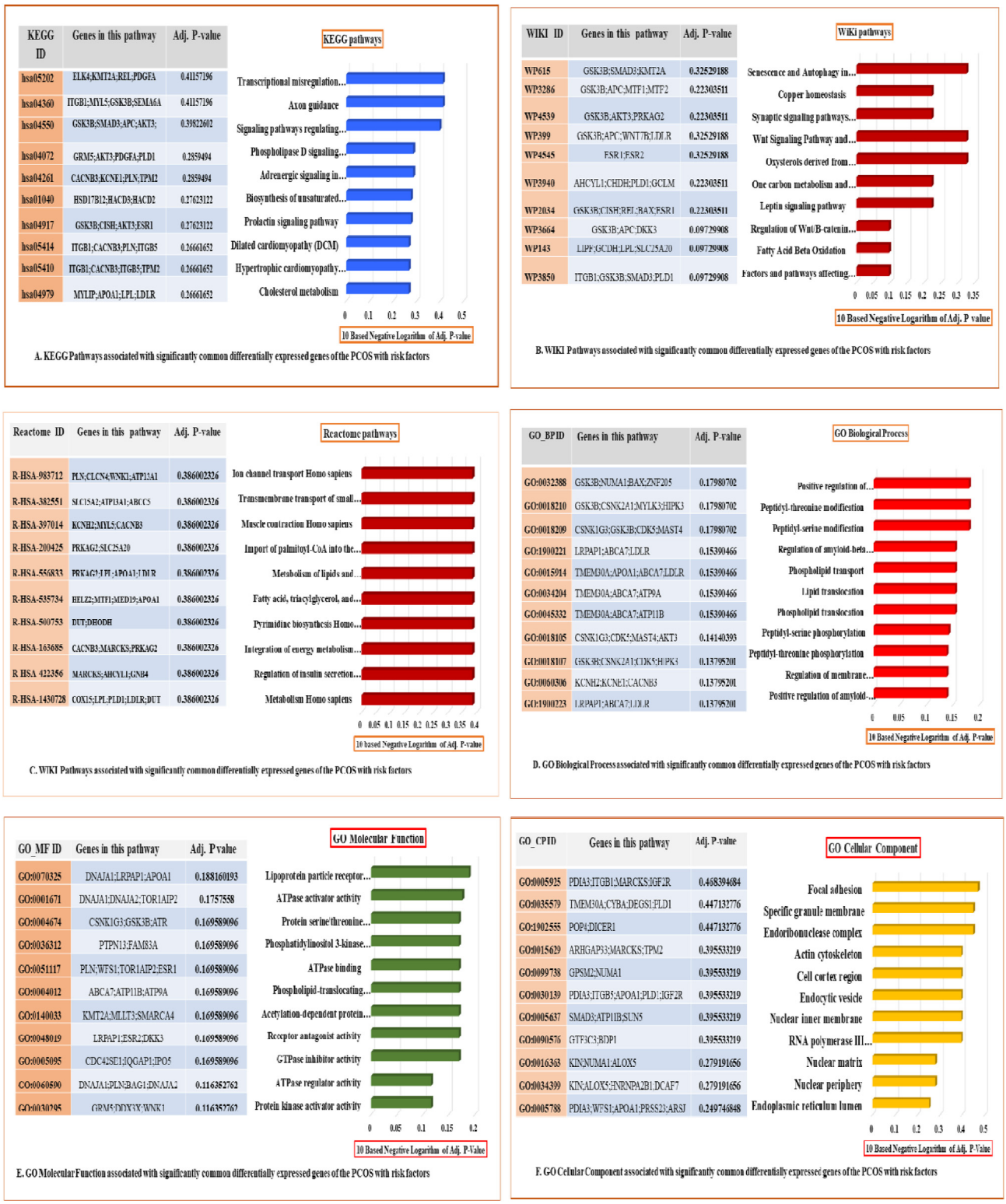


Figure 4. Pathway analysis for identifying the most significant signaling and GO pathways common to PCOS and the metabolic comorbidities revealed by the paired common differentially expressed genes. These include significant signaling pathways, and Gene Ontology is identified by (A) KEGG pathway, (B) WiKi pathway, (C) Reactome pathway, (D) GO Biological process, (E) GO Molecular Function, and (F) GO Cellular Component. Top-10 terms from each datasets were selected based on the number of genes involved and P-value less than or equal to 0.05.

receiving gonadotropin-releasing hormone (GnRH) analog/recombinant human (rh) follicle stimulating hormone (FSH) therapy for in vitro fertilization (IVF). For type 2 diabetes samples, omentum (visceral adipose) tissue were collected from both healthy control and diabetic subject as compared to age and BMI matched normal glucose tolerant women.

For respect to obesity, control samples were collected from second trimester fetus of lean mother in gestational age: 16 + 5 compared to case in gestational age: 15 + 6. Platelets from patient with acute MI within 6 h of the onset of symptoms were selected for CVD as a control samples. Blood donors with different age ranges such as 28–45, 45–60 and >60

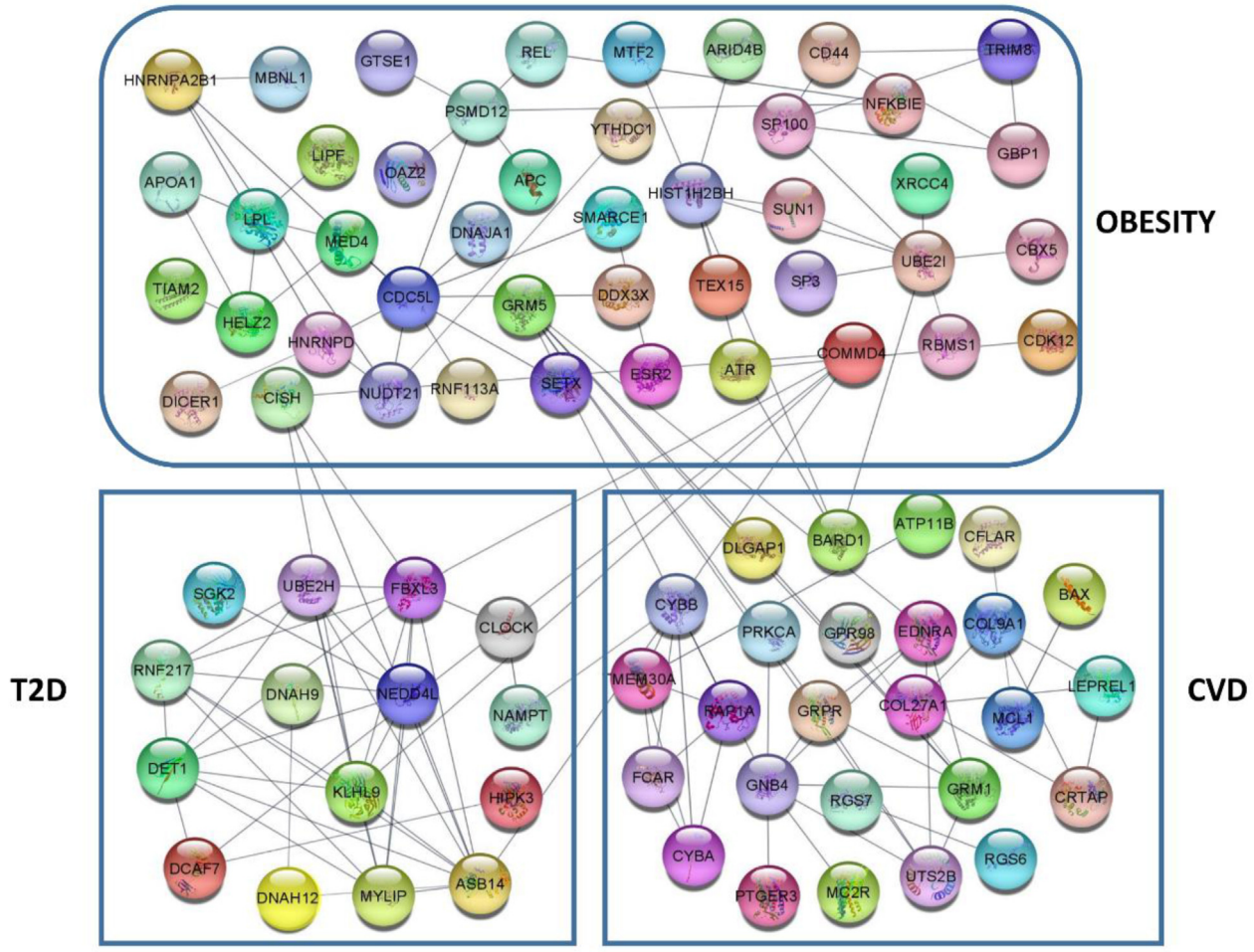


Figure 5. Protein-Protein interaction network for the common paired DEGs between PCOS and T2D (type 2 diabetes), OBS (obesity), and CVD (cardiovascular disease). The clusters indicates the 3 metabolic diseases for which they are dysregulated between PCOS.

were excluded from any cardiovascular sufferance by appropriate clinical tests.

After selection of data, we analyzed and identified 2224, 381, 600 and 1937 significant DEGs ($P < 0.05$ and log FC > 1.5) in PCOS, T2D, Obesity and CVD respectively. Among them, the number of upregulated genes were 1315, 294, 307 and 1289, while the number of down-regulated genes were 909, 132, 293 and 648, respectively. We carried out a cross-comparative analysis of the gene expression profiles to comprehend the pathogenetic involvement of PCOS with the aforementioned disorders. The Venn diagram of shows that PCOS shared 91, 205, and 34 dysregulated genes between Obesity, CVD, and T2D [Figure 2(A–C)].

To visualize their association, a gene–disease relationship network (GDN) was constructed centered on the PCOS as shown in Figure 3 (A, B). Particularly, five genes (i.e. YTHDC1, ARID4B, CCDC71L, OVOL2 and CYBB) were downregulated among PCOS, obesity and CVD, while two genes SYN2 and ESYT3 were common among PCOS, T2D and CVD. Likewise, downregulation of DPH6, ICA1, ZNF7 and ESR2 were observed in PCOS, T2D and Obesity. Conversely, PCOS shares upregulation of eleven genes (i.e. ZFYVE16, LONRF2A, ZFYVE1 etc.) with Obesity and CVD. The most critical DEGs are defined using our proposed method, summarized in Table 1.

3.2. Pathway and functional enrichment analysis

A wide variety of signaling pathways and GO terms are involved in the orchestration and development of diseases in complex disorders. In this process, we identified key pathways and gene ontologies that may connect PCOS and obesity, CVD and T2D disorders using paired common 260, 351, and 98 DEGs. Figure 4(A–F) shows 30 significant pathways and 30 GO terms from these datasets. Most of the common pathways are related to metabolism including cholesterol metabolism, Hypertrophic and dilated cardiomyopathy (DCM), prolactin signaling pathway and biosynthesis of unsaturated fatty acids from KEGG. Besides, positive regulation of amyloid-beta clearance, regulation of membrane repolarization, peptidyl-threonine-serine phosphorylation and phospholipid translocation are the most common significant GO pathway identified by GO biological process.

3.3. Protein-protein interaction network and hub protein identification

A PPI network has been built from the common shared DEGs' interactions, which consists of 85 nodes and 137 edges (Figure 5). Each node in the network is a protein, and the edge is a link between two proteins. The diagram depicts the involvement and association of signature genes in the PPI network. From the standpoint of PPI, we can also examine the association between the disorders. The network is also divided into three clusters, each of which is a metabolic disease. Several associated nodes in a PPI network are designated as hub genes at the same time.

We employed maximum centrality clique (MCC) algorithm to determine the hub-proteins and identified the top-10 hub-nodes (i.e. IGF2R, GNB4, UBE21, ASB14, CDC5L, FBXL3, DET1, NUDT4, NED4L, and KLHL9) within the PPI network. For MCC methods, top-10 hub-nodes are ranked with red to yellow-colored gradient (Figure 6). Table 2 shown the

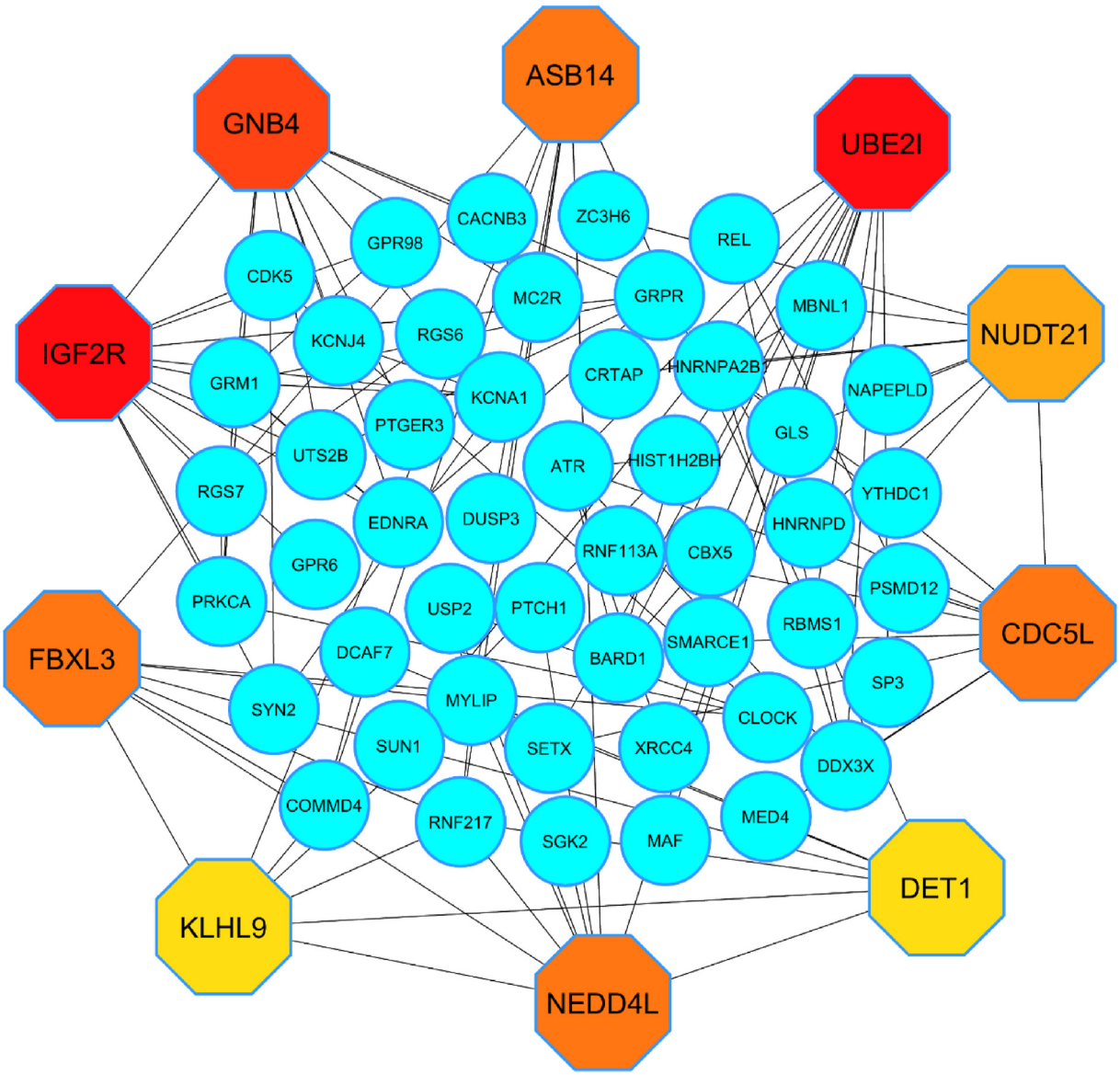


Figure 6. Ten hub proteins with octagonal shaped identified from protein-protein interaction network analysis where ranked with red to yellow-colored gradient. Round-shaped nodes indicate the most interconnected proteins among 10 hub proteins.

molecular function of identified 10 hub nodes. In each method, our selected hub genes were confined to provide higher interaction score and IGF2R showing a greater score. We have also provided the hub score of all cytoHubba plugins in Supplementary File S1. These hub nodes can be used as a potential therapeutic target for these diseases.

3.4. *In silico technology*

3.4.1. *Selection of compounds and potential drugs against IGF2R protein*

Another strategy for identifying a lead in the drug development process is the virtual scanning of molecular docking. We took ten proteins in our study with their importance through cytoHubba plugin analysis. The hub protein, IGF2R, is chosen for molecular docking, IGF2R is a most significant hub protein according to maximal clique centrality (MCC) algorithm in cytoHubba plugin. From literature review, it was indicated that IGF2R linked to PCOS among Obesity, T2D and CVD compare to others hub proteins [7, 68, 69]. It is also a type-I transmembrane glycoprotein that contains a wide N-terminal area, a single membrane region, and a short cytoplasmic tail. The sequence analysis shows that 15 homologous domains are in an extracellular area [70]. Responsive mouse model studies have demonstrated that IGF2R ligands are active in embryonic growth, placental development, signaling pathways, and tumor suppression [71].

A total of 26 compounds have been selected for IGF2R protein molecular docking. Then we selected 19 drug compounds from the DSigDB database (Enrichr tool) [48] as a control group [(Figure 7(A, B)]. Based on binding affinity, five compounds including, Kaempferol (−9.1 kcal mol^{-1}), Quercetin (−9.0 kcal mol^{-1}), Coumestrol (−8.9 kcal mol^{-1}), Sesamin (−8.7 kcal mol^{-1}), and Enzastaurin (−12.5 kcal mol^{-1}) considered for further analysis.

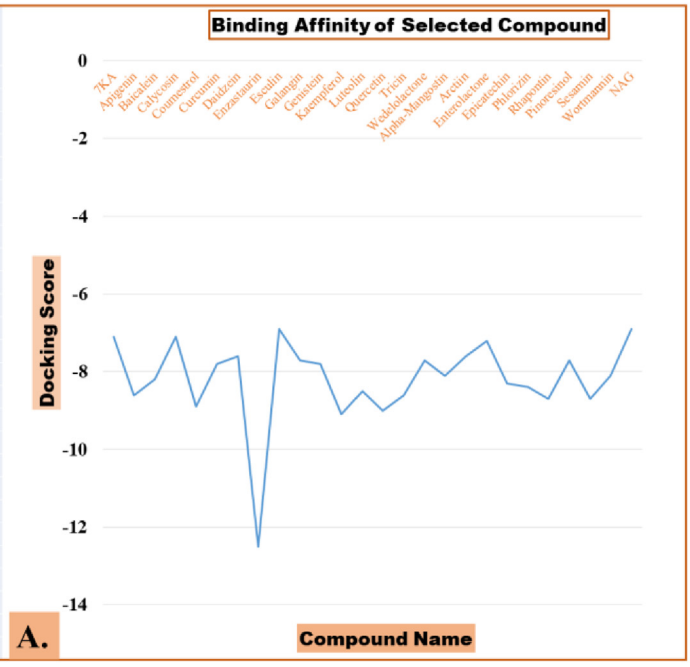
Enzastaurin is a targeted, investigative oral agent drug used to treat brain cancer, lymphoma, lung cancer, colorectal cancer, and pancreatic cancer (Drug Bank ID: DB06486).

We have selected 10 PDB IDs for molecular docking purposes but found only 2V5O with a high docking score. Besides, this PDB ID contains four human IGF2R domains that can be a significant part of molecular docking. In addition, most of the compounds bind to the active site of this domain region. NAG (2-acetamido-2-deoxy-beta-D-glucopyranose) is an inhibitor of this PDB ID, and we have used NAG as a control ligand that is already bound to the IGF2R protein. In a docking approaches, control ligand which

Table 2. Molecular function of hub proteins.

Gene Name	Protein Name	Molecular Function	UniProt ID
IGF2R	Insulin-like growth factor type 2 receptor	Transports phosphorylated lysosomal enzymes to lysosomes from the Golgi complex and the cell surface.	P11717
GNB4	Guanine nucleotide-binding protein subunit beta-4	Engaged in different transmembrane signalling systems as a modulator or transducer.	Q9HAV0
ASB14	Ankyrin repeat and SOCS box protein 14	Can be a SCF-like ECS substrate-recognition component (Elongin-Cullin–SOCS–box protein) of E3 ubiquitin-protein ligase complex.	A6NK59
CDC5L	Cell division cycle 5-like protein	Acts as a transcription activator by involving in cell cycle regulation.	Q99459
FBXL3	F-box/LRR-repeat protein 3	Involved in circadian rhythm function.	Q9UKT7
DET1	DET1 homolog	Essential for ubiquitination and later degradation of target proteins	Q7L5Y6
NUDT4	Diphosphoinositol polyphosphate phosphohydrolase 2	It can have a role in transducing the signal	Q9NZJ9
NED4L	E3 ubiquitin-protein ligase NEDD4-lik	Shown in the TOR signal regulation	Q96PU5

Compound Name	Binding Affinity(kcal/mol)	CID
7KA	-7.1	70674592
Apigenin	-8.6	5280443
Baicalein	-8.2	5281605
Calycosin	-7.1	5280448
Coumestrol	-8.9	5281707
Curcumin	-7.8	969516
Daidzein	-7.6	5281708
Enzastaurin	-12.5	176167
Esculin	-6.9	5281417
Galangin	-7.7	5281616
Genistein	-7.8	5280961
Kaempferol	-9.1	5280863
Luteolin	-8.5	5280445
Quercetin	-9	5280343
Tricin	-8.6	5281702
Wedelolactone	-7.7	5281813
Alpha-Mangostin	-8.1	5281650
Arctiin	-7.6	100528
Enterolactone	-7.2	10685477
Epicatechin	-8.3	1203
Phlorizin	-8.4	6072
Rhapontin	-8.7	637213
Pinoresinol	-7.7	73399
Sesamin	-8.7	72307
Wortmannin	-8.1	312145
NAG	-6.9	195591



Compound Name	Binding Affinity(kcal/mol)	CID
2-Methylcholine	-3.6	16942
Acetaminophen	-5.9	1983
Aspirin	-5.4	2244
Atrazine	-5.1	2256
Azacyclonol	-7.1	15723
Bisulfite	-3.5	104748
Captopril	-4.8	44093
Cephaeline	-7.8	442195
Cyperquat	-6.5	39484
Etifenin	-6	170344
Genistein	-8.2	5280961
Glycidamide	-4	91550
Lobeline	-7.4	101616
L-proline	-4.4	145742
Meclofenoxate	-5.6	4039
Menadione	-5.9	4055
Metformin	-4.9	4091
Resveratrol	-7.3	445154
Emetine	-8.2	10219

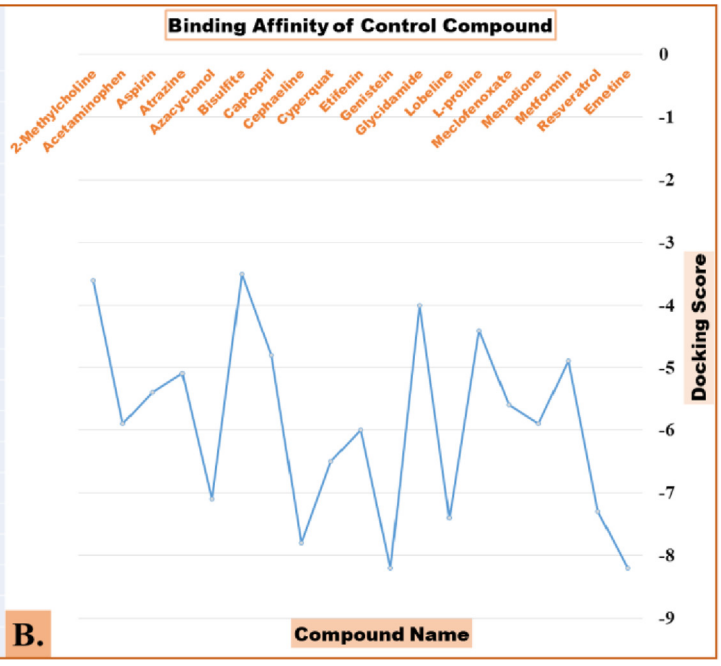


Figure 7. Docking scores of all compounds. (A) 25 phytochemicals and Enzastaurin drug and (B) 19 drug compounds of DsigDB database as a control group.

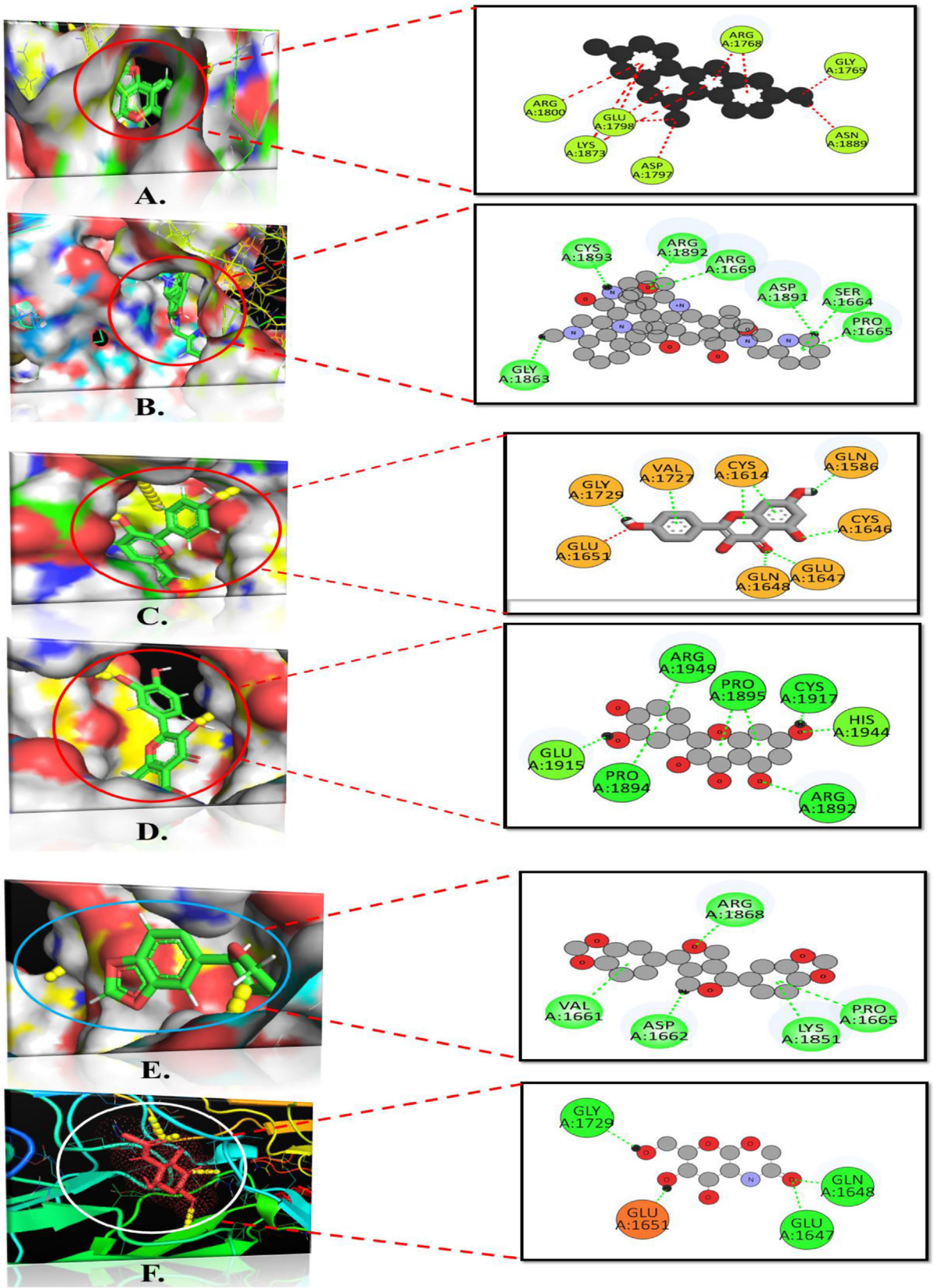


Figure 8. Molecular interactions analysis of selected compounds against IGF2R (PDB ID: 2V5O). (A) Coumestrol, (B) Enzastaurin, (C) Kaempferol, (D) Quercetin, (E) Sesamin, and (F) native ligand NAG. All selected compounds interact with the vital substrate management catalytic site.

are always bound to protein can validate our result. For this purposes, 2-acetamido-2-deoxy-beta-D-glucopyranose (NAG) can be act as a control ligand which were already bound to 2V5O protein 3D structure. NAG as a native ligand known as a potent inhibitor (binding affinity −6.9 kcal/mol) against IGF2R protein. N-Acetylglucosamine (NAG) participates biochemically in synthesizing glycoproteins known as proteoglycans in the extracellular matrix of the connective tissue (Drug Bank ID: DB00141).

3.4.2. Molecular interaction analysis of selected compounds

To illustrate fold conservation, the domain 11–14 structure covers the domain 13 (fibronectin type-II insertion) of the human protein, which is the most significant domain part of IGF2R protein [57]. The structural domain includes 15 canonical repeating extracellular units of about 147 AA, each involving four Disulfide bonds. The most conserved area of this repeat is a range of 13 AA with cysteines at both ends, and domain 13 contains an 1896 to 1944 amino acid sequence identified by the Interpro server [72].

The CASTp server predicted the active site of IGF2R (PDB ID: 2V5O) (i. e;:GLN1586, Glu 1651, GLU1647, ASP1662, SER1664, PRO1665, ARG1669, GLY1729, VAL1727, ARG1768, GLY1769, ASP1797, GLU1798, GLY1863, LYS1873, ASP1891, ARG1892, CYS1893, PRO1894, PRO1895, CYS1917, HIS1944, and ARG1949) [62]. In our analysis, the interaction of docked compounds showed that Kaempferol interacts with most of Cys 1646, Gln 1648, Glu 1651, and Val 1727 catalytic residues, which indicated that native ligand NAG (2-acetamido-2-deoxy-beta-D-glucopyranose) also bind the same catalytic site. While other compounds like Enzastaurin binds to the GLY1863, ASP1891, ARG1892, and CYS1893 amino acid residues through hydrogen bonding and SER1664, PRO1665, ARG1669 residues through hydrophobic bonding in the protein IGF2R. The BIOVIA discovery studio visualized the ligand interaction with substrate-binding pocket residue shown in Figure 8(A-F). Coumestrol, Enzastaurin, Kaempferol, Quercetin, Sesamin, and NAG showed 4, 4, 6, 3, 2, and 3 hydrogen bonds presented in Table 3 with detailed information.

3.5. 3 analysis of ADMET properties

Table 4 shows ADMET characteristics, including physicochemical, lipophilicity, water-solubility, pharmacokinetics, medicinal, and toxicity of the five selected compounds. Regarding the drug-likeness activity, all compounds do not violate Lipinski's rule of five. Both compounds display a negative response to skin sensitization and the hepatotoxicity properties of a predictable drug molecule. In the discussion section, details were explained.

3.6. Data validation of the findings

The biomarker genes found by our pathway have been cross-validated with the gold standard Benchmark database: OMIM and dbGAP. As shown in Figure 9, genes linked with T2D, Obesity, and CVD are likely to positively co-related with PCOS. Overall, these results take massive gaps in our knowledge of PCOS pathology and could lead to new directions for mechanistic correlations among PCOS and related metabolic comorbidities in female patients.

4. Discussion

Network-based approaches provide a more robust framework and provide a systematic overview of the molecular function in disease progression. These studies can reveal new links between PCOS and other metabolic disorders such as type 2 diabetes, obesity, and cardiovascular disease.

PCOS is a complex illness with many different clinical presentations that are primarily caused by hormonal deficiencies and have serious health and financial consequences. The current study compiled a list of crucial hub genes, pathways, and genes involved in PCOS and metabolic disorder pathogenesis using microarray data analysis. Obesity, CVD, and T2D all shared a total of 91, 205, and 34 dysregulated genes with PCOS. We identified several essential pathways related to PCOS, T2D, obesity, and CVD in this study, including the prolactin and leptin signaling pathways, dilated and hypertrophic cardiomyopathy, insulin secretion control, and oxysterols derived from cholesterol, etc. In contrast, Glintborg D (2010) and Ojeda-Ojeda M et al. reported that PCOS signaling pathways are primarily related to inflammation and immune with the synthesis of chemokines and interleukins [73, 74].

Peptidyl threonine and serine modification and phosphorylation, ATPase activator and regulator activity, protein serine/threonine kinase activity, protein kinase activator activity, actin cytoskeleton, focal adhesion, and receptor antagonist activity, all of which are closely linked to the progression of metabolic disorder and PCOS, which were reported in our GO study. GTF3C3 and BDP1 were found to be involved in actin

Table 3. Binding affinity results of 6 compounds with score energy and interaction with amino acids.

Compound Name	Docking energy (kcal.mol^{-1})	Hydrogen Bond Ligand atom-Amino Acid Distance (Å)	Hydrophobic Bond Interaction-Amino Acid Distance(Å)
Coumestrol	−8.9	:UNK0:H - A:ASN1889:OD1 (2.27) A:GLY1769:HN -:UNK0:O (1.90) A:GLU1798:HN -:UNK0:O (2.21) A:ASP1797:HA -:UNK0:O (2.82)	Pi-Sigma- A:ARG1768:HD2 -:UNK0 (2.41) Pi-Alkyl -:UNK0 - A:LYS1873(5.48) Pi-Alkyl-:UNK0 - A:ARG1800 (5.32)
Enzastaurin	−12.5	:UNK0:H - A:CYS1893:O (2.25) A:ARG1892:HD1 -:UNK0:O (2.35) :UNK0:H19 - A:ASP1891:OD1 (2.79) :UNK0:H13 - A:GLY1863:O (3.00)	Pi-orbitals-A:SER1664:C,O; PRO1665:N -:UNK0(4.97) Pi-Alkyl-:UNK0 - A:ARG1669 (4.97) Pi-Alkyl-:UNK0 - A:PRO1665 (4.04)
Kaempferol	−9.1	A:GLU1647:HN -:UNK0:O (2.27) A:GLN1648:HN -:UNK0:O (1.86) :UNK0:H - A:GLY1729:O (2.25) :UNK0:H -:GLU1651:O (2.21) :UNK0:H - A:GLN1586:O (2.66) A:CYS1646:HA -:UNK0:O (2.42)	Pi-Alkyl-:UNK0 - A:CYS1614(4.91) Pi-Alkyl-:UNK0 - A:VAL1727(4.34) Electrostatic-A:CYS1917(4.21)
Quercetin	−9	A:HIS1944:HE2 -:UNK0:O (2.15) :UNK0:H - A:GLU1915:OE1 (2.23) A:ARG1892:HD1 -:UNK0:O (2.79)	Pi-cation-A:ARG1949:NH2 -:UNK0(4.28) Pi-Alkyl-:UNK0 - A:PRO1895 (5.49) Pi-Alkyl-:UNK0 - A:PRO1894 (5.38)
Sesamin	−8.7	A:ARG1868:HE -:UNK0:O (2.31) :UNK0:H05 - A:ASP1662:OD2 (2.81)	Pi-Alkyl-:UNK0 - A:PRO1665 (5.27) Pi-Alkyl-:UNK0 - A:LYS1851(5.01) Pi-Alkyl-:UNK0 - A:VAL1661(5.29)
NAG (**2-acetamido-2-deoxy-beta-D-glucopyranose**)	−6.9	:UNK0:H - A:GLY1729:O (2.24) A:GLU1647:HN -:UNK0:O (2.51) A:GLN1648:HN -:UNK0:O (1.87)	Pi-Alkyl-:UNK0:H - A:GLY1729:O (5.01)

cytoskeleton regulation in our research. At the same time, Haoran Shen et al. reported that ACTB and MYH9 in module 1 were up-regulated DEGs in PCOS and were positively linked to actin cytoskeleton regulation [38]. However, two separate studies indicated that the two genes ACTB and MYH9 could lead to false positives result from PCOS-supported abnormal expression of cytoskeletal proteins [75, 76].

In this analysis, we developed a PPI network based on STRING databases using common shared differential genes. Using topological approaches, these discovered hub proteins can be employed to identify prospective therapeutic targets or biomarkers. In the PPI network analysis, ten hub proteins were identified (i.e. IGF2R, GNB4, UBE21, ASB14, CDC5L, FBXL3, DET1, NUDT4, NED4L, and KLHL9) are involved in the PCOS and metabolic disease progression. However, Haoran Shen et al. carried out their analysis on other PCOS microarray data and found that LYN, ARHGAP4, ARHGAP9, and RHOG are promising candidate genes in PCOS development [38].

The hub proteins we have identified in this study play a significant role in PCOS and metabolic disease. However, the molecular mechanism of PCOS and its link to the metabolic disease is controversial. However, several proteins are studied in a therapeutic way to identify genes that are highly crucial to our disease progression. Then we selected IGF2R protein according to topological MCC metrics ranking and literature review for in silico based studies, including virtual screening, molecular docking, and ADMET analysis with some phytochemicals compound and several drug compound from the DSigDB database as a comparing group. This study also investigating the insulin-like growth factor type 2 (IGF2R) genetic connection to PCOS, Obesity, T2DM and CVD [7, 68, 69].

Molecular simulation has been part of the current drug network, particularly for docking studies [77]. Moreover, time and laboratory costs can be significantly minimized by choosing a computational-based drug design [78]. However, of the 26 phytochemicals assessed, the peak five are selected based on their highest score, including Coumestrol, Enzastaurin, Kaempferol, Quercetin, and Sesamin are suggested to be used as promising IGF2R protein inhibitors that are closely connected to the positive control inhibitor NAG. We also concluded surprisingly that a control group of 19 drug compounds displayed a lower binding affinity than our chosen plant products and Enzastaurin drug.

Enzastaurin is an oral serine-threonine kinase inhibitor and investigational, targeted drug for the treatment of relapsed glioblastoma multiforme (GBM), an active and malignant type of brain cancer via the PKC-B and PI3K/AKT pathways, which is aimed at more than 100 sites around the world (Drug Bank Id: DB06486). M. M. Faul et al. and H. Mellor (1998) demonstrated that specifically, Enzastaurine binds to the ATP binding site inside the catalytic domain of protein kinase C (PKC), thus

Table 4. ADMET analysis of top 4 compounds with the native ligand.

ADMET Properties		Sesamin	Kaempferol	Quercetin	Coumestrol	NAG
Physicochemical properties	MW (g/mol)	354.35	286.24	302.24	268.22	221.21
	NHA	26	21	22	20	15
	NAHA	12	16	16	17	0
	HBA	6	6	7	5	6
	HBD	0	4	5	2	5
	MR	90	76.01	78.03	73.81	47.19
	TPSA (A2)	55.38	111.13	131.36	83.81	119.25
Lipophilicity	iLOGP	3.46	1.7	1.63	1.8	0.2
	XLOGP3	2.68	1.9	1.54	2.76	−1.16
	WLOGP	2.57	2.28	1.99	3.1	−3.08
	MLOGP	1.98	−0.03	−0.56	1.76	−2.61
	Silicos-IT Log P	3.25	2.03	1.54	2.88	−2.27
	Consensus Log	2.79	1.58	1.23	2.46	−1.78
Water Solubility	ESOL Log S	−3.93	−3.31	−3.16	−3.87	−0.28
	ESOL Class	Soluble	Soluble	Soluble	Soluble	Very Soluble
	ALI Log S	−3.5	−3.86	−3.91	−4.18	−0.85
	ALI Class	Soluble	Soluble	Soluble	Moderately soluble	Very Soluble
	Silicos-IT Log S	−4.6	−3.82	−3.24	−5.03	1.64
	Silicos-IT Class	Moderately soluble	Soluble	Soluble	Moderately soluble	Soluble
Pharmacokinetics Properties	GI absorption	High	High	High	High	Low
	CYP1A2 inhibitor	No	Yes	Yes	Yes	No
	CYP2C9 inhibitor	No	No	No	No	No
	CYP2D6 inhibitor	Yes	Yes	Yes	Yes	No
	CYP3A4 inhibitor	Yes	Yes	Yes	No	No
	Log Kp (skin permeation)	−6.56 cm/s	−6.7 cm/s	−7.05 cm/s	−5.98 cm/s	−8.47 cm/s
Drug-likeness	Lipinski	Yes; 0 violation	Yes; 0 violation	Yes; 0 violation	Yes; 0 violation	Yes; 0 violation
	Ghose	Yes	Yes	Yes	Yes	No; 1 violation
	Egan	Yes	Yes	Yes	Yes	Yes
	Bioavailability Score	0.55	0.55	0.55	0.55	0.55
Medicinal Chemistry	Pains	0 alert	0 alert	1 alert	0 alert	0 alert
	Lead-likeness	No; 1 violation: MW > 350	Yes	Yes	Yes	No; 1 violation
	Synthetic accessibility	4.12	3.14	3.23	3.16	3.79
Toxicity	Ames test	Yes	No	No	Yes	No
	hERG I inhibitor	No	No	No	No	No
	hERG II inhibitor	No	No	No	No	No
	Skin Sensitisation	No	No	No	No	No
	Hepatotoxicity	No	No	No	No	No

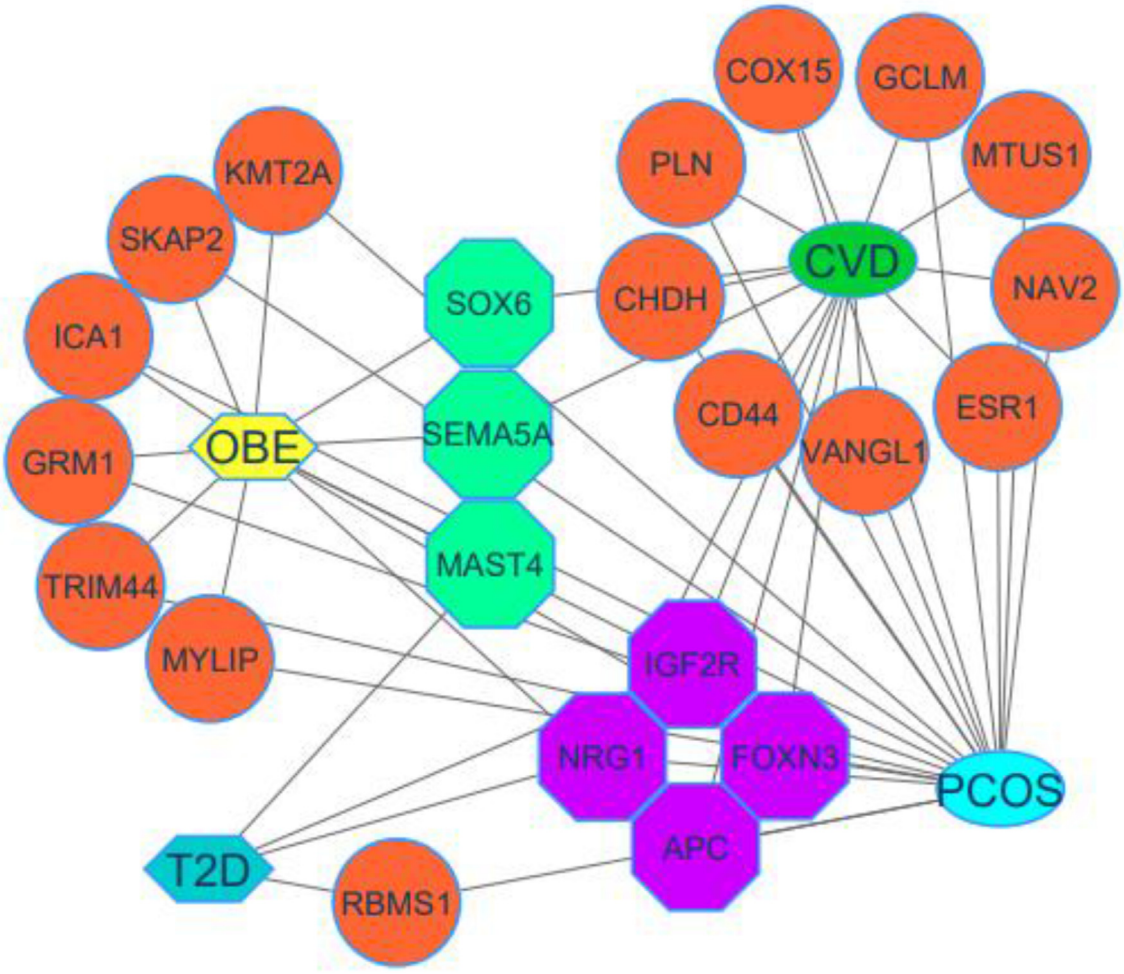


Figure 9. Gene-disease association analysis of DEGs of metabolic risk factors with PCOS using OMIM and dbGAP databases. Ellipse and hexagonal-shaped nodes represent risk factors, while round and octagonal-shaped nodes represent DEGs. Octagonal-shaped DEGs indicate common genes between PCOS and three metabolic comorbidities.

blocking downstream signaling [79,80]. In our study Enzastaurin binds to the domain number 11 and 12 by forming four hydrogen bonds at Gly1863, Asp1891, Cys1893 and Arg1892 residues with three hydrophobic bonds at Ser1664, Pro1665 and Arg1669 residues.

Flavonoids have many pharmacological impacts, including the regulative effects of anti-oxidative, anti-inflammatory, and lipid metabolism [81]. A recent study concluded that the diet of high-kaempferol foods could decrease the incidence of several diseases such as inflammation induce metabolic disorder and type 2 diabetes mellitus [82], and cardiovascular diseases [83]. Kaempferol binds to the catalytic site by forming six hydrogen bonds at Gln1586, Cys1646, Glu1647, Gln1648, Glu1651, and Gly1729, while two hydrophobic bonds at Cys1614 and Val1727 residues. Only Kaempferol binds to the same active site of positive control ligands NAG.

S. Jahan et al. reported that the effects of Quercetin on lipid profile, hormonal indicators, and glucose absorption were beneficial and can also be used in ongoing treatment as a PCOS remedial [84], which supported our findings. Our results have shown that Quercetin is a potent flavonoid capable of combating metabolic and endocrine comorbidities in PCOS. It binds to the Fibronectin type 2 domain by forming two hydrogen bonds at Glu1915 His1944 residues and one electrostatic bond at Cys1917 residue.

Another phytoestrogen with anti-diabetic, antioxidant, and anti-inflammatory effects is Coumestrol. According to H. Zywno et al. studies, Coumestrol could treat obesity and type 2 diabetes through nutritional therapy [85]. According to our findings, coumestrol binds to the catalytic site by forming two hydrophobic bonds at Lys1873 and Asn1889 residues and one electrostatic bond at Arg1768 residue. These two studies suggested that Sesamin is a novel therapy for type 2 diabetes patients with cardiovascular dysfunction like coronary artery disease, nephropathy, and atherosclerosis [86, 87]. Natural antioxidants, such as Sesamin, have been shown in a recent study to effectively prevent the severity of Covid-19 [72]. Our findings indicated that Sesamin actively binds to the Arg1868 and Asp1662 residues, forming hydrogen bond interactions.

Fortunately, all of the compounds bind to the active site of the protein, as indicated by the CASTp server, and the domain parts 11 and 12, except for Quercetin, which only binds to the domain number of 13 (Fibronectin type 2) of the IGF2R protein. Several studies confirm the functionality of this domain. The probable IGF–II–binding site, according to Schmidt et al. and Brown et al., is located in domain 11, the first in the IGF2R extracellular region identified by X-ray crystallography [81, 82]. While domain 13 enhances IGF-II binding, which is thought to be mediated by FNII [63], J. Linnell et al. discovered that this decreases the rate of IGF-II release [83].

ADMET analysis was used in a productive and cost-efficient manner to assess all compounds' physicochemical and drug-like properties [88]. Computational biology technology has improved the drug development process in the biopharmaceutical sector to identify and establish lead compounds [89]. ADMET evaluation offers a good indication of

prospective candidates for medications. Potential drugs are expected to have an optimal molecular weight between 150 and 500 g/mol (dalton), the higher number of heavy atoms (NHA) and higher number of heavy aromatic atoms (NAHA), hydrogen bonding acceptors (HBA) $\leq$ 10, hydrogen bonding donor (HBD) $\leq$ 5, TPSA (topological polar surface area) around 20 to 130 Å2, and molar refractory (MR) range of 40–130 which considered a good sign of drug development [66, 90]. Analysis of water solubility indicated that all compounds are soluble in water and maximum showed a positive response in CYP inhibitor properties. The medicinal chemistry properties of four compounds show that they have no PAINS warning undoubtedly to achieve their targets with no harmful aspects. In addition, all compounds have lead-like properties except Sesamin and NAG. Tests for human ether go-go-genes (hERG) have revealed that these compounds have not been inhibited. The results of the Ames test also show that only coumestrol and sesamin compounds are mutagenic compared to others.

The goal of systems biology-based network studies is to identify human diseases by utilizing genome-scale biological networks, high-throughput experimental datasets, and topological analytic methods [87]. The AMPK and adipocytokine signaling pathways are involved in the development of metabolic problems in PCOS women, according to the molecular foundation of AR and STK11 genes [91]. Although the system biology and network based technique have been applied in breast cancer, B cell lymphoma, colorectal cancer, prostate cancer, research employing this technique to investigate the relationships between PCOS and endometrial cancer is sparse [91, 92]. Low levels of high-density lipoprotein cholesterol (HDL-C) and low-density lipoprotein cholesterol (LDL-C) are all common dyslipidaemia patterns linked with polycystic ovarian syndrome [93]. Numerous study shown that PCOS strongly link to the risk of metabolic disorders but system biology and network based approaches will not be performed yet to identify DEGs of PCOS and metabolic abnormalities. Besides, several similar studies have been established to identify potential biomarkers and pathways that reveal the molecular pathogenesis of type 2 diabetes, tuberculosis and rheumatoid arthritis [94]. COVID-19 [40] and systemic sclerosis with cancers [95].

In this investigation, we employed system biology and network-based analysis against IGF2R hub proteins, we have also examined potential treatment by the 29 phytochemicals through molecular docking and ADMET properties analysis. In addition, we propose some latent targets for PCOS and comorbidities which could help develop new therapeutic techniques. So, we investigate the transcriptomic data for identifying DEGs as the molecular target with therapeutic agents. Aside from that, further investigation into the function of these active compounds is required, which will aid in the development of new disease-specific drugs.

5. Conclusions

The microarray transcriptomic data were used in this research to determine the genetic association of PCOS with a metabolic comorbidity. Our results suggest that the positive relationship between PCOS and T2D, obesity, and CVD can be identified in these network-based techniques. Moreover, we have established several molecular and gene ontology pathways which provide an understandable view of the genetic connection between metabolism and PCOS. The gene expression study has also identified ten hub genes, most of which contribute significantly to PCOS and metabolic disease progression. The top three compounds including Enzastaurin, Kaempferol, and Quercetin could play a role in treating PCOS and metabolic disorders as a therapeutic objective. This approach will enable us to determine how their associations may result in a better mechanistic understanding of the emergence of metabolic diseases in PCOS and propose new methods of precise gene-marker diagnosis and successful phytochemical treatment of diseases. Further in vitro study will be required to confirm genetic expression and the relation between PCOS and metabolic comorbidities and treat the progression of the disease by our predicted compounds.

Declarations

Author contribution statement

Abdullah Al Emran and Md. Arif Khan: Conceived and designed the experiments and Wrote the paper.

Md. Arju Hossain: Performed the experiments and Wrote the paper.

Sheikh Abdullah Al Ashik and Moshiur Rahman Mahin: Analyzed and interpreted the data.

Md. Al Amin: Contributed reagents, materials, analysis tools or data.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability statement

Data included in article/supp. material/referenced in article.

Declaration of interest's statement

The authors declare no conflict of interest.

Additional information

Supplementary content related to this article has been published online at https://doi.org/10.1016/j.heliyon.2022.e12480.

References

[1] T. H, D. A, M. L, Polycystic ovary syndrome: a complex condition with psychological, reproductive and metabolic manifestations that impacts on health across the lifespan, BMC Med. 8 (2010).
[2] R. Azziz, D.A. Dumesic, M.O. Goodarzi, Polycystic ovary syndrome: an ancient disorder? Fertil. Steril. 95 (5) (2011) 1544–1548.
[3] C. McCartney, J. Marshall, Polycystic ovary syndrome. Author manuscript, N. Engl. J. Med. 375 (1) (2016) 54–64.
[4] L. Doretto, F.C. Mari, A.C. Chaves, Polycystic ovary syndrome and psychotic disorder, Front. Psychiatr. 11 (2020) 1–6.
[5] S. Franks, et al., Development of polycystic ovary syndrome: involvement of genetic and environmental factors, Int. J. Androl. 29 (1) (2006).
[6] F. Wang, et al., Expression and clinical significance of the HIF-1a/ET-2 signaling pathway during the development and treatment of polycystic ovary syndrome, J. Mol. Histol. 46 (2) (2015).
[7] S. Kaur, K.J. Archer, M.G. Devi, A. Kriplani, J.F. Strauss, R. Singh, Differential gene expression in granulosa cells from polycystic ovary syndrome patients with and without insulin resistance: identification of susceptibility gene sets through network analysis, J. Clin. Endocrinol. Metab. 97 (10) (2012).
[8] B. Xu, Y.W. Zhang, X.H. Tong, Y.S. Liu, Characterization of microRNA profile in human cumulus granulosa cells: identification of microRNAs that regulate Notch signaling and are associated with PCOS, Mol. Cell. Endocrinol. 404 (2015).
[9] R.A. Wild, Polycystic ovary syndrome: a risk for coronary artery disease? Am. J. Obstet. Gynecol. 186 (1) (2002).
[10] P. Zimmet, K.G.M.M. Alberti, J. Shaw, Global and societal implications of the diabetes epidemic, Nature 414 (6865) (2001).
[11] I.R. Ilie, C.E. Georgescu, Polycystic ovary syndrome-epigenetic mechanisms and aberrant MicroRNA, Adv. Clin. Chem. 71 (2015).
[12] S. Bednarska, A. Siejka, The pathogenesis and treatment of polycystic ovary syndrome: what's new? Adv. Clin. Exp. Med. 26 (2) (2017).
[13] J. Rojas, et al., Polycystic ovary syndrome, insulin resistance, and obesity: navigating the pathophysiologic labyrinth, Int. J. Reprod. Med. 2014 (2014).
[14] H.J. Teede, S. Hutchison, S. Zoungas, C. Meyer, Insulin resistance, the metabolic syndrome, diabetes, and cardiovascular disease risk in women with PCOS, Endocrine 30 (1) (2006).
[15] B.G. Chittenden, G. Fullerton, A. Maheshwari, S. Bhattacharya, Polycystic ovary syndrome and the risk of gynaecological cancer: a systematic review, Reprod. Biomed. Online 19 (3) (2009).
[16] J.M. Schildkraut, P.J. Schwingl, E. Bastos, A. Evanoff, C. Hughes, Epithelial ovarian cancer risk among women with polycystic ovary syndrome, Obstet. Gynecol. 88 (4) (1996).
[17] B.C.J.M. Fauser, et al., Consensus on women's health aspects of polycystic ovary syndrome (PCOS): the Amsterdam ESHRE/ASRM-Sponsored 3rd PCOS Consensus Workshop Group, Fertil. Steril. 97 (1) (2012).
[18] B.K. Tripathi, A.K. Srivastava, Diabetes mellitus: Complications and therapeutics 12, 2006, pp. 130–147, no. 7.

[19] L. Chen, D.J. Magliano, P.Z. Zimmet, Mellitus — present and future perspectives, Nat. Publ. Gr. 8 (4) (2011) 228–236.
[20] A. Gastaldelli, Role of beta-cell dysfunction , ectopic fat accumulation and insulin resistance in the pathogenesis of type 2 diabetes mellitus, Diabetes Res. Clin. Pract. 93 (2011) S60–S65.
[21] V. Hljl, Pathogenesis and Glycemic Management of Type 2 Diabetes 15, 2012, pp. 239–246, no. 4.
[22] Bogari NM. Genetic construction between polycystic ovarian syndrome and type 2 diabetes. Saudi J. Biol. Sci.. 2020 Oct 1;27(10):2539-2543.
[23] E.O. Talbott, J.V. Zborowski, J.R. Rager, K.E. Kip, X. Xu, T.J. Orchard, Polycystic ovarian syndrome (PCOS): a significant contributor to the overall burden of type 2 diabetes in women, J. Wom. Health 16 (2) (2007).
[24] S. Weerakiet, C. Srisombut, P. Bunnag, S. Sangtong, N. Chuangsoongnoen, A. Rojanasakul, Prevalence of type 2 diabetes mellitus and impaired glucose tolerance in Asian women with polycystic ovary syndrome, Int. J. Gynecol. Obstet. 75 (2) (2001).
[25] S.F.A. Grant, et al., Confers risk of type 2 diabetes 38, 2006, pp. 320–323, no. 3.
[26] E. Zeggini, M.I. Mccarthy, TCF7L2: the biggest story in diabetes genetics since HLA ?, 2007, pp. 1–4.
[27] C.J. Groves, et al., Replication and Confirms TCF7L2 as a Type 2 Diabetes Susceptibility Gene with a Substantial Effect on 55, 2006.
[28] M.S. Mohammed, S. Sendra, J. Lloret, I. Bosch, Review Article Systems and WBANs for Controlling Obesity 2018, 2018.
[29] J. Vrbikova, V. Hainer, Obesity and Polycystic Ovary Syndrome, 2009, pp. 26–35.
[30] F. Escobar-morreale, A. Francisco, L.S. Milla, The polycystic ovary syndrome associated with morbid obesity may resolve after weight loss induced by Bariatric surgery 90, 2005, pp. 6364–6369, no. 12.
[31] A.D.A. Statements, The Impact of Prevention on Reducing the Burden of Cardiovascular Disease, 2008.
[32] P. Pinola, et al., Normo- and hyperandrogenic women with polycystic ovary syndrome exhibit an adverse metabolic pro fi le through life, 2017.
[33] J. Schmidt, K. Landin-wilhelmsen, M. Bra, E. Dahlgren, Cardiovascular disease and risk factors in PCOS women of postmenopausal age : a 21-year, Controlled Follow-Up Study 96 (2014) 3794–3803.
[34] A. Woodward, D. Broom, C. Dalton, M. Metwally, M. Klonizakis, Supervised exercise training and increased physical activity to reduce cardiovascular disease risk in women with polycystic ovary syndrome: study protocolS for a randomized controlled feasibility trial, Trials 21 (1) (2020) 1–8.
[35] W. Koenig, M. Karakas, C. Herder, J. Baumert, B. Thorand, Oxidized LDL and the risk of coronary heart disease : results from the METHODS, Clin. Chem. 57 (8) (2011) 1196–1200.
[36] H.Y. Liu, J.Q. Liu, Z.X. Mai, Y.T. Zeng, A subpathway-based method of drug reposition for polycystic ovary syndrome, Reprod. Sci. 22 (4) (2015).
[37] H.Y. Liu, Y.L. Huang, J.Q. Liu, Q. Huang, Transcription factor-microRNA synergistic regulatory network revealing the mechanism of Polycystic ovary syndrome, Mol. Med. Rep. 13 (5) (2016).
[38] H. Shen, Z. Liang, S. Zheng, X. Li, Pathway and network-based analysis of genome-wide association studies and RT-PCR validation in polycystic ovary syndrome, Int. J. Mol. Med. 40 (5) (2017).
[39] G. Wu, E. Dawson, A. Duong, R. Haw, L. Stein, ReactomeFIViz: a Cytoscape app for pathway and network-based data analysis [v2 ; ref status: indexed, analysis], 2014, pp. 1–14.
[40] M.I. Hasan, M.H. Rahman, M.B. Islam, M.Z. Islam, M.A. Hossain, M.A. Moni, Systems Biology and Bioinformatics approach to Identify blood based signatures molecules and drug targets of patient with COVID-19, Inform. Med. Unlocked 28 (2022), 100840.
[41] M.H. Rahman, et al., Bioinformatics and system biology approaches to identify pathophysiological impact of COVID-19 to the progression and severity of neurological diseases, Comput. Biol. Med. (2021) 104859.
[42] M.H. Rahman, et al., A network-based bioinformatics approach to identify molecular biomarkers for type 2 diabetes that are linked to the progression of neurological diseases, Int. J. Environ. Res. Publ. Health 17 (3) (2020).
[43] B. Rahmani, E. Ghashghayi, M. Zendehdel, A. Baghbanzadeh, M. Khodadadi, Molecular mechanisms highlighting the potential role of COVID-19 in the development of neurodegenerative diseases, Phys. Int. 109 (2) (2022) 135–162.
[44] J.C. Oliveros, VENNY. An Interactive Tool for Comparing Lists with Venn Diagrams. http//bioinfogp.cnb.csic.es/tools/venny/index.html, 2007.
[45] A. Jia, L. Xu, Y. Wang, Venn diagrams in bioinformatics, Briefings Bioinf. 22 (5) (2021).
[46] T.A. Taz, et al., Network-based identification genetic effect of SARS-CoV-2 infections to Idiopathic pulmonary fibrosis (IPF) patients, Briefings Bioinf. 22 (2) (2021) 1254–1266.
[47] M.E. Smoot, K. Ono, J. Ruscheinski, P.L. Wang, T. Ideker, Cytoscape 2.8: new features for data integration and network visualization, Bioinformatics (2011).
[48] M. V Kuleshov, et al., Enrichr: a comprehensive gene set enrichment analysis web server 2016 update, Nucleic Acids Res. 44 (W1) (2016). W90–W97.
[49] D. Szklarczyk, et al., The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible, Nucleic Acids Res. (2017).
[50] C.H. Chin, S.H. Chen, H.H. Wu, C.W. Ho, M.T. Ko, C.Y. Lin, cytoHubba: identifying hub objects and sub-networks from complex interactome, BMC Syst. Biol. (2014).
[51] C.Y. Li, J.H. Cai, J.J.P. Tsai, C.C.N. Wang, Identification of hub genes associated with development of head and neck squamous cell carcinoma by integrated bioinformatics analysis, Front. Oncol. 10 (2020).
[52] B. Mi, et al., A network pharmacology study on analgesic mechanism of Yuanhu-Baizhi herb pair, J. Pharm. Biomed. Anal. 189 (14) (2020) 113475.
[53] Q. Yang, K. Li, X. Li, J. Liu, Identification of key genes and pathways in myeloma side population cells by bioinformatics analysis, Int. J. Med. Sci. 17 (14) (2020) 2063–2076.
[54] K. Wang, et al., Revealing the mechanisms and the material basis of Rubia cordifolia L. on abnormal uterine bleeding with uniting simultaneous determination of four components and systematic pharmacology approach-experimental validation, J. Pharm. Biomed. Anal. 189 (2020) 113475.
[55] M.G. Aucar, C.N. Cavasotto, Molecular docking using quantum mechanical-based methods, Methods Mol. Biol. 2114 (2020) 269–284.
[56] S. Saikia, M. Bordoloi, Molecular docking: challenges, advances and its use in drug discovery perspective, Curr. Drug Targets 20 (5) (2019) 501–521.
[57] J. Brown, et al., Structure and functional analysis of the IGF-II/IGF2R interaction, EMBO J. 27 (1) (2008) 265–276.
[58] S.K. Burley, et al., RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences, Nucleic Acids Res. (2021).
[59] D. Studio, Dassault Systemes BIOVIA, Discovery Studio Modelling Environment, Release 4.5, Accelrys Softw. Inc., 2015.
[60] L.L.C. Schrödinger, The PyMOL Molecular Graphics System, Version 2.4, Schrödinger LLC, 2020.
[61] W. Kaplan, T.G. Littlejohn, Swiss-PDB viewer (deep view), Briefings Bioinf. (2001).
[62] W. Tian, C. Chen, X. Lei, J. Zhao, J. Liang, CASTp 3.0: computed atlas of surface topography of proteins, Nucleic Acids Res. 46 (W1) (2018) W363–W367.
[63] S. Kim, et al., PubChem 2019 update: improved access to chemical data, Nucleic Acids Res. (2019).
[64] S. Dallakyan, A.J. Olson, Small-molecule library screening by docking with PyRx, Methods Mol. Biol. (2015).
[65] T.A. Halgren, Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94, J. Comput. Chem. (1996).
[66] A. Daina, O. Michielin, V. Zoete, SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules, Sci. Rep. (2017).
[67] D.E. V Pires, T.L. Blundell, D.B. Ascher, pkCSM: predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures, J. Med. Chem. 58 (9) (2015) 4066–4072.
[68] Y.-J. Choi, et al., Effect of prenatal bisphenol A exposure on early childhood body mass index through epigenetic influence on the insulin-like growth factor 2 receptor (IGF2R) gene, Environ. Int. 143 (2020), 105929.
[69] C.-C. Feng, et al., Cardiac apoptosis induced under high glucose condition involves activation of IGF2R signaling in H9c2 cardiomyoblasts and streptozotocin-induced diabetic rat hearts, Biomed. Pharmacother. 97 (2018) 880–885.
[70] P. Lobel, N.M. Dahms, S. Kornfeld, Cloning and sequence analysis of the cation-independent mannose 6-phosphate receptor, J. Biol. Chem. 263 (5) (1988).
[71] P. Ghosh, N.M. Dahms, S. Kornfeld, Mannose 6-phosphate receptors: new twists in the tale, Nat. Rev. Mol. Cell Biol. 4 (3) (2003).
[72] A. Mitchell, et al., The InterPro protein families database: the classification resource after 15 years, Nucleic Acids Res. 43 (D1) (2015) D213–D221.
[73] D. Glintborg, M. Andersen, An update on the pathogenesis, inflammation, and metabolism in hirsutism and polycystic ovary syndrome, Gynecol. Endocrinol. 26 (4) (2010).
[74] M. Ojeda-Ojeda, M. Murri, M. Insenser, H. Escobar-Morreale, Mediators of low-grade chronic inflammation in polycystic ovary syndrome (PCOS), Curr. Pharmaceut. Des. 19 (32) (2013).
[75] N.R. Salvetti, E.J. Gimeno, J.A. Lorente, H.H. Ortega, Expression of cytoskeletal proteins in the follicular wall of induced ovarian cysts, Cells Tissues Organs 178 (2) (2004).
[76] M. Cortón, et al., Differential gene expression profile in omental adipose tissue in women with polycystic ovary syndrome, J. Clin. Endocrinol. Metab. 92 (1) (2007).
[77] A.R. Oany, A.-A. Emran, T.P. Jyoti, Design of an epitope-based peptide vaccine against spike protein of human coronavirus: an in silico approach, Drug Des. Dev. Ther. 8 (2014) 1139–1149.
[78] T. Talele, S. Khedkar, A. Rigby, Successful applications of computer aided drug discovery: moving drugs from concept to the clinic, Curr. Top. Med. Chem. (2010).
[79] M.M. Faul, et al., Acyclic N-(azacycloalkyl)bisindolylmaleimides: isozyme selective inhibitors of PKCβ, Bioorg. Med. Chem. Lett 13 (11) (2003).
[80] H. Mellor, P.J. Parker, The extended protein kinase C superfamily, Biochem. J. 332 (2) (1998).
[81] Y. Fang, Y.C. Xu, M.M. An, F. Wang, Y.Y. Jiang, Research advances in pharmacokinetics and pharmacological effects of flavonoids, Pharmaceut. Care Res. 15 (1) (2015).
[82] C. Luo, et al., Kaempferol alleviates insulin resistance via hepatic IKK/NF-κB signal in type 2 diabetic rats, Int. Immunopharm. 28 (1) (2015).
[83] J. Lin, et al., Dietary intakes of flavonols and flavones and coronary heart disease in US women, Am. J. Epidemiol. 165 (11) (2007).
[84] S. Jahan, et al., Therapeutic potentials of Quercetin in management of polycystic ovarian syndrome using Letrozole induced rat model: a histological and a biochemical study, J. Ovarian Res. 11 (1) (2018).
[85] H. Zywno, et al., The influence of coumestrol on sphingolipid signaling pathway and insulin resistance development in primary rat hepatocytes, Biomolecules 11 (2) (2021).
[86] T.D. Thuy, et al., Novel therapeutic effects of sesamin on diabetes-induced cardiac dysfunction, Mol. Med. Rep. 15 (5) (2017).
[87] O.O. Oguntibeju, Type 2 diabetes mellitus, oxidative stress and inflammation: examining the links, Int. J. Physiol. Pathophysiol. Pharmacol. 11 (3) (2019).

[88] A. Rahman, N.H. Naheed, S.C. Raka, N. Qais, A.Z.M.R. Momen, Ligand-based virtual screening, consensus molecular docking, multi-target analysis and comprehensive ADMET profiling and MD stimulation to find out noteworthy tyrosine kinase inhibitor with better efficacy and accuracy, Orient. Pharm. Exp. Med. (2019).
[89] A.S. Ivanov, A.V. Veselovsky, A.V. Dubanov, V.S. Skvortsov, Bioinformatics platform development: from gene to lead compound, Methods Mol. Biol. (2006).
[90] C.A. Lipinski, F. Lombardo, B.W. Dominy, P.J. Feeney, Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, Adv. Drug Delivery Rev. (2012).
[91] P. Gollapalli, N.S. Kumari, P. Shetty, T.S. Gnanasekaran, Molecular basis of AR and STK11 genes associated pathogenesis via AMPK pathway and adipocytokine signalling pathway in the development of metabolic disorders in PCOS women, Beni-Suef Univ. J. Basic Appl. Sci. 11 (1) (2022).
[92] J.L. Griffin, J.P. Shockcor, Metabolic profiles of cancer cells, Nat. Rev. Cancer 4 (7.) (2004).
[93] D.A. Ehrmann, et al., Prevalence and predictors of the metabolic syndrome in women with polycystic ovary syndrome, J. Clin. Endocrinol. Metab. 91 (1) (2006).
[94] M.I. Hasan, M.A. Hossain, P. Bhuiyan, M.S. Miah, M.H. Rahman, A system biology approach to determine therapeutic targets by identifying molecular mechanisms and key pathways for type 2 diabetes that are linked to the development of tuberculosis and rheumatoid arthritis, Life Sci. 297 (2022), 120483.
[95] M.K. Islam, et al., Network based systems biology approach to identify diseasome and comorbidity associations of Systemic Sclerosis with cancers, Heliyon 8 (2) (2022).