Identification of key genes and molecular pathways in type 2 diabetes mellitus and polycystic ovary syndrome via bioinformatics analyses

J. ZHANG¹, F.-J. ZHANG², L. ZHANG³, D.-X. XIAN², S.-A. WANG⁴, M. PENG⁵, Y. LIU⁶

¹Department of Pediatrics, Shandong Second Provincial General Hospital, Jinan, China
²College of Acupuncture and Massage, Shandong University of Traditional Chinese Medicine, Jinan, China
³Department of Clinical Pharmacy, Shaoxing People’s Hospital, Shaoxing Hospital, Zhejiang University School of Medicine, Shaoxing, China
⁴Department of Pediatric Surgery, Shandong Provincial Hospital affiliated to Shandong First Medical University, Jinan, China
⁵Department of Traditional Chinese Medicine, Shandong Provincial Hospital affiliated to Shandong First Medical University, Jinan, China

Jing Zhang, Fengjun Zhang and Lin Zhang contributed equally to this work

Abstract. – OBJECTIVE: Type 2 diabetes mellitus (T2DM) and polycystic ovary syndrome (PCOS) are highly prevalent endocrine system diseases. However, studies on the molecular mechanisms of T2DM and PCOS at the transcriptomic level are still few. Thus, we aimed to reveal the potential common genetic and molecular pathways between T2DM and PCOS via bioinformatics analyses.

MATERIALS AND METHODS: We downloaded the GSE10946 and GSE18732 datasets for T2DM and PCOS, respectively, from the National Center for Biotechnology Information’s Gene Expression Omnibus (GEO) database. These datasets were subjected to integrated differential and weighted gene co-expression network analyses (WGCNA) to screen common genes. Thereafter, functional enrichment and disease gene association analyses were performed, transcription factor (TF)-gene and TF-miRNA-gene regulatory networks were constructed, and finally, the relevant target drugs were identified.

RESULTS: We identified common genes (BIRC3, DEPTOR, TNNL3, ADRA2A) in T2DM and PCOS. Pathway enrichment analysis depicted that the common genes were enriched in smooth muscle contraction, channel inhibitor activity, apoptosis, and tumor necrosis factor (TNF) signaling pathways. TFs such as SP7, KLF8, HCFC1, IRF1, and MLLT1 played key roles in TF regulatory networks. Orlistat was indicated to be an important gene-targeting drug.

CONCLUSIONS: This study is the first study to explore four diagnostic biomarkers and gene regulatory networks for T2DM and PCOS. The findings of our study provide novel insights into the diagnosis and treatment of T2DM and PCOS.

Key Words: Type 2 diabetes mellitus (T2DM), Polycystic ovary syndrome (PCOS), Endocrine system diseases, Molecular mechanisms, Gene regulatory networks, Bioinformatics.

Introduction

Type 2 diabetes mellitus (T2DM) is a heterogeneous disease characterized by high blood glucose levels and accounts for approximately 90-95% of adults with diabetes¹. The number of people with this disease is expected to reach 642 million worldwide by 2040², imposing a severe economic burden on patients and their families. Typical clinical symptoms of patients with T2DM include polyuria, polydipsia, polyphagia, and the life-threatening hyperosmolar hyperglycemic state. Moreover, several critical life-threatening complications such as cardiovascular disease, stroke, diabetic retinopathy,
neuropathy, and nephropathy may develop later\(^5\), with high morbidity, disability, and mortality rates. There is evidence of gender differences in T2DM and its associated complications, and such differences may be related to abnormal sex hormone secretion\(^4\). The pathogenesis of T2DM can be broadly attributed to insulin resistance and the corresponding inability of the pancreas to maintain proper insulin secretion to compensate for the decreased insulin sensitivity\(^5\). Insulin resistance is a common feature of both T2DM and polycystic ovary syndrome (PCOS). Cohort studies\(^6\) have been conducted to show that PCOS is a high-risk factor for T2DM. Through meta-analysis, Galazis et al\(^7\) have identified eight protein biomarkers, namely pyruvate kinase M1/M2, apolipoprotein A-I, albumin, peroxiredoxin 2, membrane-linked protein A2, a-1-B-glycoprotein, flotillin-1, and haptoglobin, which differed significantly between T2DM and PCOS patients. These biomarkers can aid improving our understanding of the link between T2DM and PCOS. However, studies of both these diseases at the transcriptomic level are still inadequate.

PCOS, a common gynecological disorder, is characterized by endocrine dysfunction and manifested by hyperandrogenemia, anovulation, and polycystic ovaries\(^8\). PCOS has a high prevalence among women of reproductive age, affecting about 5-10\% of this group\(^9\). The exact etiology of the disease remains unknown, and the prevailing view is that it is a familial genetic syndrome caused by a combination of environmental and genetic factors\(^10\). In the last decade, the complex mechanisms of PCOS pathogenesis have been understood mainly through multi-omics studies\(^11\), including genomics and proteomics. Among these, genomics studies\(^12\) have exhibited the relevance between insulin resistance, vitamin D, cytokines, adipokines, and the incidence of PCOS. Furthermore, inflammation plays a key role in the pathogenesis of PCOS, and elevated levels of inflammatory markers are strongly associated with PCOS development. Nevertheless, studies\(^13\) specific to the molecular pathway mechanisms of PCOS remain inadequate. PCOS is a serious threat to female reproductive health; as an endocrine disorder, it can cause metabolic syndromes such as obesity, insulin resistance, hypertension, and dyslipidemia. In pathological states, patients with PCOS exhibit insulin resistance, insulin hypofunction, and dysregulation of glucose homeostasis in several tissues. Insulin resistance is considered to be the underlying cause of metabolic syndromes and is closely associated with T2DM and PCOS\(^14\). This evidence implies some association between PCOS and T2DM; however, this association is vague and unsystematic.

With the development of gene chip and sequencing technologies, thousands of genetic data expressed in various diseases can be screened rapidly, thus contributing to a deeper understanding of the pathogenesis of diseases at the genetic level. In this study, we aimed to identify common genes associated with T2DM and PCOS. To this end, we planned to subject the sample datasets of T2DM and PCOS from the National Center for Biotechnology Information’s Gene Expression Omnibus (GEO) database to weighted gene co-expression network analysis (WGCNA) and differential analysis. Further, we intended to perform enrichment analysis, gene regulatory network construction, disease association analysis, and gene target drugs discovery for the common genes to explore the common pathogenic mechanisms between T2DM and PCOS, thereby providing ideas for diagnosing and treating the diseases.

**Materials and Methods**

**Raw Data Download and Processed**

GSE18732 and GSE10946 datasets were downloaded from the GEO database using the keywords “T2DM” and “PCOS” respectively. The T2DM dataset is hosted on the GPL9486 platform (Affymetrix Human Genome U133-Plus-2.0 Array), containing 45 T2DM patients and 47 control samples, whereas the PCOS dataset is hosted on the GPL570 platform (Affymetrix Human Genome U133-Plus-2.0 Array), containing 12 samples of PCOS patients and 11 control samples\(^15\). We used the Perl software (available at: https://www.perl.org) script to perform ID transformation, missing value processing, and removal of batch effects on the matrix files uploaded by the contributors and the corresponding platform annotation files. This was done to remove probe groups without the corresponding gene symbols and genes with more than one probe group, respectively, and to force the normalization of the data. Finally, a gene matrix with row names as sample names and column names as gene symbols was obtained and used for subsequent analysis.
Bioinformatics studies on T2DM and PCOS

Identification of Differentially Expressed Genes (DEGs) for T2DM and PCOS

GEO2R is an online microarray data analysis tool developed by the GEO database based on 2 R packages (available at: https://www.r-project.org), GEO query and Limma (Bioconductor, Roswell Park Comprehensive Cancer Center, NY, USA). The GEO query package is used to read the data, while the Limma package is used to calculate the multiples of differential expressions. We employed GEO2R to compare the gene expression profiles between different groups of the two datasets for T2DM and PCOS to identify DEGs between the disease and control groups. Thresholds for the DEGs in the T2DM and PCOS datasets were set to $p < 0.05$ and $|\log_2 FC| \geq 0.5$. The ggplot2 package of the R software (available at: https://www.r-project.org) was used to produce heat maps and gradient volcano plots.

Weighted Gene Co-Expression Network Analysis (WGCNA)

WGCNA is an algorithm that analyzes gene co-expression patterns of multiple samples, identifies biologically significant co-expressed gene modules, and explores the relationship between gene networks and disease\textsuperscript{16}. Here, we used WGCNA to obtain PCOS- and T2DM-associated modules. Before the analysis, we performed a hierarchical cluster analysis using the Hclust function in R language to exclude abnormal samples. Thereafter, we used the “pick Soft Threshold” function in the WGCNA package (available at: https://www.r-project.org) to select a suitable soft threshold $\beta$ value (ranging from 1-25) according to the criteria of scale-free networks. Subsequently, the soft threshold $\beta$ values between all gene pairs and the gene correlation matrix were calculated using Pearson’s correlation coefficient analysis to construct the adjacency matrix. The topological overlap matrix and corresponding phase dissimilarity were transformed from the adjacency matrix. Further, a hierarchical clustering dendrogram was constructed, and similar gene expressions were classified into different modules. Finally, the expression profile of each module was summarized by the module signature gene, and the correlation between Module eigengene (ME) and clinical features was calculated. Generally, the WGCNA algorithm is suitable for processing complex transcriptomic data. However, according to the official WGCNA guidelines\textsuperscript{16}, the number of samples in the dataset needs to be greater than 15 to be considered meaningful for the data obtained. Thus, the genes that could not be clustered were uniformly classified as gray modules, which are considered meaningless gene modules and excluded.

In this study, we used the Venn tool in Omicshare (available at: https://www.omicshare.com/tools/Home/Soft/venn) to obtain crossover genes for the DEGs and WGCNA module genes of T2DM and PCOS. The resulting common genes were used for the next step of molecular pathway identification and gene network construction.

Gene Enrichment Analysis of T2DM and PCOS

Gene enrichment analysis is an enrichment molecular pathway analysis of typical genes to obtain annotated information on genes regarding molecular function and chromosomal localization. Gene Ontology (GO) is employed for annotating biological pathways, including biological processes (BP), molecular functions (MF), and cellular components (CC). Kyoto Encyclopedia of Genes and Genomes (KEGG) is used in the study of metabolic pathways and plays a crucial role in gene annotation and is widely used in bioinformatics for pathway studies.

Based on Ensemble\textsubscript{104} or 51 versions, species selection for humans (GRCh38.p13) and type selection for genes, eventually, GO terms and KEGG pathways were obtained through Omicshare. Values with $p < 0.05$ were considered statistically significant.

Transcription Factors (TFs) Regulatory Networks Construction

Transcription factors (TFs) are protein molecules that can bind to specific genes, thereby ensuring that the target gene is expressed at a specific intensity in a specific time and space, which is crucial in molecular studies. Here, we employed the NetworkAnalyst platform (available at: https://www.networkanalyst.ca/) to screen reliable TFs from the JASPAR database\textsuperscript{17} with common genes to construct a TF-gene regulatory network. miRNA is a small non-coding RNA that has been shown\textsuperscript{18} to regulate gene expression by promoting mRNA degradation or inhibiting mRNA translation. Therefore, in this study, we further explored the pattern of miRNAs and TFs co-regulating T2DM and PCOS. We established the TF-miRNA-gene regulatory network and obtained the relevant information from the RegNetwork repository.
Gene-Disease Association Analysis

With the intensive study of transcriptomics, gene regulation-related disease profiles are continuously identified, and different diseases can be linked by one or more similar genes, and this plays an important role in the process of disease comorbidity identification and prevention. DisGeNET is a comprehensive database for the study of disease comorbidity and disease genetic features, with data integrated from multiple information sources, including the genome-wide association studies (GWAS) catalog and literature collection sources, featuring different biomedical manifestations of the disease and highlighting new insights into human genetic disorders. We conducted a study related to gene-disease comorbidity through the NetworkAnalyst 3.0 platform (available at: https://www.networkanalyst.ca, equipped with the DisGeNET database) to identify comorbid diseases related to T2DM and PCOS.

Identification of Drug Candidates

Identifying gene-drug interactions was a vital part of the study. The Drug Signatures Database (DSigDB) is an archive for identifying gene-associated targeted drugs. The database has 22,527 gene sets, and access is obtained through the Enrichr platform (available at: https://maayanlab.cloud/Enrichr/), an open-source online enrichment analysis tool with a large number of different gene set libraries for gene enrichment information and drug information queries. Based on the common genes of T2DM and PCOS, the DSigDB equipped with the Enrichr platform identified relevant drugs and small molecule compounds, providing new ideas for the prevention and treatment of T2DM and PCOS.

Statistical Analysis

Differential analysis and WGCNA were conducted using R software version 4.1.0 (available at: https://www.r-project.org) to identify gene expression patterns in disease and control groups. The hypergeometric test was employed for enrichment analysis, while paired data comparisons were performed using Wilcoxon test. Detailed statistical methods employed in processing transcriptome data are described in the Materials and Methods section. In this study, $p < 0.05$ was considered statistically significant.

Results

Identification of DEGs for T2DM and PCOS

We performed differential analysis on the GSE18732 and GSE10946 datasets. Based on the threshold of $p < 0.05$ and $|\log FC| \geq 0.5$, 87 (33 upregulated and 54 downregulated) DEGs were obtained for the GSE18732 dataset, whereas 151 (61 upregulated and 90 downregulated) DEGs were obtained for the GSE10946 dataset. Heatmaps were used to represent the top 80 DEGs (Figure 1A, 1C). Gradient volcano plots were displayed in Figure 1B, 1D, and the up-regulated and down-regulated genes varied with the fold change and $p$-value. The top ten DEGs were specifically labeled in the gradient volcano plots. Detailed information on the differential analysis of the GSE18732 and GSE10946 datasets were provided in Supplementary Tables I-II, respectively. The DEGs from the GSE18732 and GSE10946 datasets were crossed to obtain a common gene (ADRA2A) in Figure 2.

Analysis of Gene Co-Expression Modules in T2DM and PCOS

We performed WGCNA analysis on the GSE18732 and GSE10946 datasets. Here, the cut line for the GSE18732 dataset was set to 98, and four outlier sample sets (GSM465294, GSM405503, GSM465319, GSM465281) were excluded, whereas that of the shear line of the GSE10946 dataset was set to 65 and GSM277445 dataset was excluded. We selected 4 and 12 as the optimum soft threshold $\beta$ values for the GSE18732 and GSE10946 datasets, respectively (Figure 3A-B). The minimum number of module genes was set to 50 for both datasets. We identified 7 modules in the GSE18732 dataset for yellow, turquoise, green, red, blue, brown, and grey (Figure 3C). Similarly, 11 modules were identified in the GSE10946 dataset, with each color representing a different module, including brown, yellow, black, pink, red, magenta, purple, greenyellow, blue, green, and grey (Figure 3D). Heatmaps on module-trait relationships based on Pearson’s correlation coefficients were then plotted to assess the association of each module with the disease. The green module ($p = 3e-04$) was significantly associated with T2DM (cor = 0.38) and contained 135 genes (Figure 3E). The pink module significantly correlated with PCOS ($p < 0.05$) which was identified as a typically associated module for PCOS (cor =
Bioinformatics studies on T2DM and PCOS

We performed GO enrichment analysis of the four common genes attained in the former analyses and obtained 201 BP, 20 MF, and 8 CC (Supplementary Table III). Figure 5 demonstrated the GO terms for the top five. For 0.5), containing 146 genes (Figure 3F). Finally, 3 intersecting genes (BIRC3, DEPTOR, TNNL3) were obtained by taking the intersection of modular genes (Figure 4). Subsequently, DEGs were cross-linked with co-expression module genes to obtain common genes for T2DM and PCOS.

**Enrichment Analysis of the Common Genes for T2DM and PCOS**

We performed GO enrichment analysis of the four common genes attained in the former analyses and obtained 201 BP, 20 MF, and 8 CC (Supplementary Table III). Figure 5 demonstrated the GO terms for the top five. For

---

**Figure 1.** Identification of DEGs in T2DM and PCOS. A, Heatmap of differentially expressed genes (DEGs) in T2DM. B, Volcano map of T2DM. The red nodes represented the up-regulated DEGs and the blue nodes represented the down-regulated DEGs. C, Heatmap of DEGs in PCOS. D, Volcano map of PCOS. The red node represented the up-regulated DEGs and the blue nodes represented the down-regulated DEGs.

**Figure 2.** The intersection of DEGs in the GSE18732 and GSE10946 dataset. T2DM included 53 down-regulated DEGs, 33 up-regulated DEGs, while PCOS included 89 down-regulated DEGs, 61 up-regulated DEGs. ADRA2A was a common down-regulated gene.
BP, the top 10 GO terms were the regulation of smooth muscle contraction, smooth muscle contraction, regulation of muscle contraction, regulation of muscle system process, regulation of blood circulation, muscle contraction, negative regulation of cell size, epinephrine transport, negative regulation of insulin secretion involved in cellular response to glucose stimulus, negative regulation of hydrolase activity (Table I). The GO terms of the top ten molecular functions we-
Bioinformatics studies on T2DM and PCOS

Figure 4. The intersection of WGCNA in the GSE18732 and GSE10946 dataset. There were 132 genes in the GSE18732 and 143 genes in the GSE10946, and 3 genes were obtained as common genes by overlapping the two datasets, include BIRC3, DEPTOR, TNNL3.

Figure 5. GO analysis of T2DM and PCOS related to the common genes.

- Calcium channel inhibitor activity, thioesterase binding, catecholamine binding, adrenergic receptor binding, cysteine-type endopeptidase inhibitor activity involved in apoptotic process, ion channel inhibitor activity, channel inhibitor activity, cysteine-type endopeptidase inhibitor activity involved in apoptotic process, calcium channel regulator activity, and G protein-coupled amine receptor activity (Table I). As for CC, the most prominent GO terms were striated muscle thin filament, myofilament, sarcomere, basolateral plasma membrane, myofibril, contractile fiber, basal plasma membrane, and basal part of cell (Table I).

The KEGG signaling pathway analysis (Table II, Supplementary Table IV) revealed five signaling pathways, including apoptosis - multiple species, platinum drug resistance, small cell lung cancer, toxoplasmosis, and TNF signaling pathway.
Construction of TF Regulatory Networks for T2DM and PCOS

To identify the changes occurring at the transcriptional level in the T2DM and PCOS common genes, we used NetworkAnalyst (available at: https://www.networkanalyst.ca) to explore the TF regulatory networks. Both TF-gene and TF-miRNA-gene regulatory networks were imported into Cytoscape software (available at: https://cytoscape.org) to synthesize a visual network. Figure 6 showed that up to 65 TF co-regulated TNNI3 in the TF-gene interaction network. We found that TFs (SP7, KLF8, HCFC1, IRF1, and MLLT1) were involved in the regulation of two common genes.

The TF-miRNA-gene regulatory network was generated using NetworkAnalyst and involved two parts (Figure 7), with 28 miRNAs and 24 TF genes co-regulating the common genes. We observed that the two networks contained 58 nodes and 56 edges and that TFAP2C, TFAP2A, CTCF, USF1 co-regulated two common genes.

Identification of Gene-Disease Associations

In our analysis of gene-disease correlations using NetworkAnalyst, we found 239 diseases that were strongly associated with common genes (Supplementary Table V), of which the top ten diseases were cardiomegaly, malignant neoplasm of gallbladder, gallbladder carcinoma, other restrictive cardiomyopathy, familial restrictive cardiomyopathy (disorder), heart diseases,
Figure 6. Transcription factor-gene regulatory network in the common genes. The circular dots represented critical genes, and the octagonal dots attached next to the critical genes represent transcription factors that regulated the critical genes. In addition, the quadrilateral-shaped transcription factors indicated regulation of two critical genes, include SP7, KLF8, HCFC1, IRF1, and MLLT1.

Figure 7. Transcription factors-miRNA-gene regulatory networks in the common genes. There were two TF-miRNA networks. Pink plots represented common genes, purple plots represent TF genes, and the yellow plots represented miRNAs. Network (A) had 39 plots and 40 edges, which consisted of 3 critical genes, 23 TF genes, and 13 miRNAs. Network (B) had 17 plots and 16 edges including 1 critical gene, 1 TF genes, and 15 miRNAs.
fatigue, gastrointestinal lymphoma, thrombotic thrombocytopenic purpura, and acquired, young adult onset (Figure 8).

**Identification of Drug Candidates**

Based on the T2DM and PCOS common genes, we predicted the related drug compounds from the DSigDB through the Enrichr platform. A total of 339 related drugs were obtained. The top ten compounds were extracted based on the p-value (Table III), including diethylstilbestrol, thimerosal, 67526-95-8, diethylstilbestrol CTD 00005818, trichostatin A, cyclophosphamide, BP 897, Brimonidine, IB-MECA, Orlistat.

**Discussions**

Presently, T2DM and PCOS are highly prevalent diseases that have a large impact on human health and quality of life. There are increasing bioinformatics studies on both these diseases. Related studies20,21 have identified PIK3R1, RAC1, GNG3, GNAI1, CDC42, and ITGB1 as biomarkers of T2DM. Moreover, Bi et al22 have presented the crucial role of inflammation and immunity in the development and progression of PCOS through bioinformatic screening and validation and exhibited the potential of TLR2, TLR8, and CD14 as therapeutic targets for PCOS. However, the transcriptomic aspects of the mechanisms between these two diseases have not yet been investigated.

In this study, a bioinformatics approach was used for the first time to explore the common genes and molecular pathway features between T2DM and PCOS. We identified the DEGs and co-expression module genes from the GSE18732 and GSE10946 datasets by bioinformatics analysis and obtained four common genes (BIRC3, DEPTOR, TNNL3, ADRA2A) after taking the intersection using the Venn diagram. Subsequently, we performed functional and pathway enrichment analysis of these common genes and found that common genes were enriched in smooth muscle contraction, channel inhibitor activity, apoptosis, and tumor necrosis factor. In addition, TF-gene and TF-miRNA-gene regulatory networks were constructed with SP7, KLF8, HCFC1, IRF1 and MLLT1 playing important roles in the TF regulatory networks. Orlistat was an important gene targeting drug for T2DM and PCOS. Genetic co-morbidity analysis showed that common genes were associated with metabolism-related diseases and heart diseases.

GO enrichment analysis of T2DM and PCOS common genes showed that the biological processes of both the common genes were significantly enriched in smooth muscle contraction. Smooth muscle is mainly distributed in the arteriovenous vessel walls, bladder, uterus, and digestive, respiratory, and reproductive tracts in humans. Wang et al23 have demonstrated through in vitro rat experiments that PCOS reduces muscle contractility in the gastrointestinal tract and that

<table>
<thead>
<tr>
<th>Term</th>
<th>p-value</th>
<th>Combined score</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diethylstilbestrol</td>
<td>5.29E-05</td>
<td>3,385.06</td>
<td>TNNI3; ADRA2A</td>
</tr>
<tr>
<td>Thimerosal CTD 00006868</td>
<td>2.53E-04</td>
<td>1,275.31</td>
<td>TNNI3; BIRC3</td>
</tr>
<tr>
<td>67526-95-8 ctd 00007263</td>
<td>4.16E-04</td>
<td>468.2089</td>
<td>DEPTOR; ADRA2A; BIRC3</td>
</tr>
<tr>
<td>Diethylstilbestrol CTD 00005818</td>
<td>7.98E-04</td>
<td>610.308</td>
<td>DEPTOR; TNNI3</td>
</tr>
<tr>
<td>Trichostatin A CTD 00006660</td>
<td>0.00129782</td>
<td>451,663.8</td>
<td>DEPTOR; TNNI3; ADRA2A; BIRC3</td>
</tr>
<tr>
<td>Cyclophosphamide CTD 00005734</td>
<td>0.00150225</td>
<td>480.0267</td>
<td>TNNI3; BIRC3</td>
</tr>
<tr>
<td>BP 897 TTD 00002536</td>
<td>0.00219831</td>
<td>4,077.188</td>
<td>ADRA2A</td>
</tr>
<tr>
<td>Brimonidine CTD 0000810</td>
<td>0.00219831</td>
<td>4,077.188</td>
<td>ADRA2A</td>
</tr>
<tr>
<td>IB-MECA CTD 00003042</td>
<td>0.00219831</td>
<td>4,077.188</td>
<td>BIRC3</td>
</tr>
<tr>
<td>Orlistat TTD 00009937</td>
<td>0.00219831</td>
<td>4,077.188</td>
<td>TNNI3</td>
</tr>
</tbody>
</table>
part of the mechanism lies in decreasing the responsiveness of acetylcholine and MLC20 phosphorylation. The extent of production of nitric oxide (NO), a vasodilator, is an important marker of vascular health. The vascular system affected by insulin resistance and T2DM expresses high levels of G protein-coupled receptor kinase 2 (GRK2), which inhibits NO production, and thus leads to cardiovascular disease. In this study, genetic prediction of the comorbidity of T2DM and PCOS was found to be associated with increased risk of cardiovascular disease. Regarding molecular functions, common genes are enriched in channel inhibitor activity. Calcium channel blockers are the most common physiological cell signaling pathways, and they affect almost every aspect of cell life. Calcium channel blockers reportedly play a protective role in the brain-renal vasculature of T2DM patients. In contrast, cellular components, are mainly concentrated in the rhabdomyotic thin filaments, myofilaments, and basolateral plasma membranes.

KEGG pathway enrichment analysis revealed that common genes were mainly enriched in signaling pathways such as apoptosis and tumor necrosis factor (TNF) signaling pathway. Apoptosis is a complex biological phenomenon that involves cell contraction, chromatin condensation, DNA fragmentation between nucleosomes, and disintegration into membrane-encapsulated vesicles (apoptotic vesicles). Increasing evidence has proved that β-cell apoptosis is an important pathogenetic mechanism of T2DM. It has also been found that MSCs release apoptotic vesicles (apoVs) upon apoptosis, and apoVs induce reprogramming of macrophages at the transcriptional level through extracellular phagocytosis, restoring intracellular macrophage homeostasis and improving T2DM. Zhen et al. found that downregulation of NEAT1 or upregulation of miR-381 helps promoting PCOS granulosa cell proliferation and inhibits apoptosis by suppressing IGF1 expression. These results suggest that the NEAT1/miR-381/IGF1 axis plays a key role in the pathogenesis of PCOS and may provide a new target for treating PCOS. These results suggest that apoptosis plays an important role in the development of PCOS disease, and we can further validate the pathway subsequently through apoptosis experiments and other experiments. TNF is a key mediator and regulator of the mammalian immune response, controlling the development of the immune system, cell survival and proliferation, and regulating metabolic processes. A correlation between PCOS, T2DM, and inflammation has been reported in vivo, and inflammation markers such as C-reactive protein, TNF-α, interleukin 6, and white blood cell count are reportedly closely related. This chronic inflammation is exacerbated by obesity and hyperinsulinemia, and studies have indicated the interaction between inflammation and hyperinsulinemia, obesity, and hyperandrogenemia. This chronic inflammatory state leads to long-term complications of diabetes, including nonalcoholic fatty liver disease, retinopathy, cardiovascular disease, and nephropathy, and may underlie the association of T2DM with other diseases such as Alzheimer’s disease, polycystic ovary syndrome, gout, and rheumatoid arthritis. Activation of pro-inflammatory pathways (e.g., TNF signaling pathway) can lead to increased production of pro-inflammatory cytokines and chemotactic mediators, which induce a local or systemic pro-inflammatory state and impaired insulin signaling. Hyperglycemia in diabetic patients is thought to lead to a dysfunctional immune response that is unable to control the spread of invasive pathogens in diabetic patients. As a result, diabetic patients are known to be more susceptible to infections and the increased prevalence of T2DM will increase the incidence of infectious diseases and related complications. Therefore, the inhibition of the inflammatory response may play a role in disease progression, such as immunomodulatory therapy. In addition to the influence of inflammatory factors, PCOS has been shown to be significantly associated with immune response. Some studies have identified immunodiagnostics biomarkers (cAMP, S100A9, TLR8, and IL6R) in PCOS patients by bioinformatics and constructed disease diagnostic models based on biomarkers that may play an important role in the disease progression process. For this reason, we considered screening T2DM and PCOS biomarkers for inflammatory response-related mechanisms and further constructing disease models in future studies to provide a reference direction for disease diagnosis, treatment and prognosis.

Among the four genes obtained, BIRC3 and ADRA2A have been previously reported to have a clear association with T2DM. BIRC3, an inhibitor of apoptosis (cIAP2), belongs to the family of IAP proteins that contain the structural threshold of BIR. In the pancreas, long-term exposure to high fatty acid levels leads to lipid overload in β-cells, impairing insulin secretion from β-cells and ultimately leading to β-cell apoptosis.
Therefore, protecting pancreatic β-cells from lipid overload may be a potential therapeutic strategy for T2DM. FGF21 has been demonstrated to reduce islet cell apoptosis by downregulating lipid accumulation in islets and enhancing BIRC3 levels in vivo through AMPK-ACC and PPARδ/γ signaling. Alogliptin can protect islet cells by phosphorylating the activated TF CERB, which stabilizes blood glucose levels in vivo. The target proteins of alogliptin include BIRC3, BLC2, and IRS2, while BIRC3 acts as an inhibitor of islet cell apoptosis. However, only few studies in literature on BIRC3 targeting PCOS are currently available, which provides inspiration to further study T2DM and PCOS.

ADRA2A encodes the α2-adrenergic receptor, a G-coupled receptor that plays a regulatory role in central nervous system signaling pathways and metabolic functions. ADRA2A variants are significantly associated with T2DM. ADRA2A knockout mice showed increased insulin secretion, and ADRA2A was identified as the gene that determines T2DM through further studies of the genetic locus in diabetic rats. This finding suggests that ADRA2A receptor antagonists could be used for the targeted treatment of T2DM.

DEPTOR has been correlated with distant complications of T2DM. DEPTOR is a DEP structural domain-containing protein that interacts with mTOR; it is a natural mTOR inhibitor that plays a chief role in biological processes such as cell proliferation, survival, metabolism, and apoptosis. Studies have established that the mTOR signaling pathway is associated with obesity and related metabolic disorders and that high mTORC1 activity contributes to the progression of diabetic nephropathy and end-stage renal disease. A study found DEPTOR gene polymorphisms had a role in T2DM vascular complications, and since DEPTOR may be involved in metabolic complications of T2DM, DEPTOR and mTOR signaling pathways could be a direction for future studies on late T2DM complications.

TNNT3 mainly encodes cardiac troponin I, which inhibits actin-myosin interactions and thus mediates transverse muscle relaxation. A study on the hypertrophic cardiomyopathy (HCM) family revealed a double heterozygous mutation in TNNT3, which can eventually lead to HCM. Patients with severe acute respiratory syndrome coronavirus 2 infections in T2DM develop acute cardiovascular syndrome; thus, the correlation between TNNT3 and diabetes could be a target for further research.

Regarding disease complications, we found that patients with T2DM and PCOS are prone to comorbid cardiac disorders as a consequence of long-term metabolic abnormalities. The core etiology and main features of polycystic ovary syndrome are hyperandrogenemia and insulin resistance. Patients with PCOS have a combination of a range of metabolic abnormalities, including abdominal obesity, hyperglycemia, hypertension, and dyslipidemia, with insulin resistance playing a key role in its pathogenesis. The metabolic syndrome increases the risk of developing T2DM by 5-fold and cardiovascular disease by 2-fold.

Based on the common genes, we constructed the TF-gene and TF-miRNA-gene regulatory networks. In the network diagram, TNNI3 interacts with TFs at the highest rate. Notably, SP7, KLF8, HCFC1, MLLT1, and IRF1 were involved in the regulation of two key genes. Among them, HCFC1 was first identified in herpes simplex virus transcription. Further studies revealed that HCFC1 is a common component of the active human CpG island promoter that plays a major role in cellular transcriptional regulation. In the epigenetic regulatory mechanisms of PCOS and diabetes, altered DNA methylation can lead to disease phenotypes. This observation provides us with new ideas to investigate the regulatory role of HCFC1 in the pathogenesis of T2DM and PCOS.

As mentioned before, for treatment, immunomodulatory therapy has a greater potential. Another study has shown that gut microbes are associated with the development of T2DM, and strong evidence suggests that the interaction between the gut microbiota, the host immune system and metabolism is a key factor in the pathophysiology of obesity and T2DM. Also, the microbiota plays an important role in the reproductive endocrine system of women throughout their lives through interactions with estrogens, androgens, insulin, and other hormones, including PCOS, endometriosis, and poor pregnancy outcomes. We therefore suggest that intestinal flora dysbiosis may be a common pathogenesis for both, and therefore treatment targeting the regulation of intestinal flora may be therapeutically helpful for the disease. In the present study, we found that Orlistat is a common therapeutic agent for T2DM and PCOS. Orlistat, which is associated with T2DM and PCOS, promotes weight loss by inhibiting pancreatic and gastric lipase, thereby reducing fat absorption from the gastrointestinal tract, while also lowering blood
Bioinformatics studies on T2DM and PCOS

In a retrospective study, orlistat and metformin had similar effects in reducing body mass index, testosterone levels, and insulin levels in overweight/obese women with PCOS. Orlistat has been reported to significantly reduce testosterone levels and insulin resistance markers and improve lipid levels. Orlistat has some therapeutic value in T2DM and PCOS. Studies identifying other drugs with these two diseases are currently lacking, and their mechanism of action can be further investigated in the future to provide a reference for clinical use.

Limitations

Our study has some limitations. Since we performed secondary mining of the GEO dataset, the lack of experimental and clinical research is a shortcoming of this study. Experimental and clinical studies targeting genes and molecular pathways common to T2DM and PCOS are urgently required for further validation and application.

Conclusions

In this study, we identified four key genes BIRC3, DEPTOR, TNNI3, ADRA2A through the bioinformatic study of T2DM and PCOS-related datasets and further explored the biological processes between T2DM and PCOS. The findings of our study indicated that apoptosis and TNF signaling pathways may be the common pathogenesis of both these diseases. As a result of long-term metabolism, cardiac disorders are a common comorbidity of these two diseases. Our study findings provide relevant drugs for clinical treatment, which should be further confirmed by a large number of subsequent experimental studies to improve clinical diagnosis and treatment and reduce the occurrence of related complications.

Conflict of Interest

The Authors declare that they have no conflict of interests.

Acknowledgements

We acknowledge the GEO database for providing its platforms and contributors for uploading their meaningful datasets. Meanwhile, we thank Dr. Guangli Sun, Chief Physician, Department of Traditional Chinese Medicine, Laixi City Hospital, for her financial assistance with this study. We thank Bullet Edits Limited for the linguistic editing and proofreading of the manuscript.

References


4) Kautzky-Willer A, Harreiter J, Pacini G. Sex and Gender Differences in Risk, Pathophysiology and Complications of Type 2 Diabetes Mellitus. Endocrine Reviews 2016; 37: 278-316.


Bioinformatics studies on T2DM and PCOS


41) Xie T, So WY, Li XY, Leung PS. Fibroblast growth factor 21 protects against lipotoxicity-induced pancreatic beta-cell dysfunction via regulation of AMPK signaling and lipid metabolism. Clin Sci (Lond) 2019; 133: 2029-2044.


