ABSTRACT

Diabetic nephropathy (DN) is the main cause of end-stage renal disease (ESRD). Bioinformatics technology was adopted to screen and analyze the core genes of early DN to explore its pathogenesis. GSE30528, GSE96804, and GSE30122 chip data were obtained from Gene Expression Omnibus (GEO) database to screen DN and healthy controls for differentially expressed genes. Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) for functional annotation and signaling pathway enrichment; String and Cytoscape were used for protein-protein interaction (PPI) network construction and core gene screening, NCBI-Genes to search the expression profile of core genes. 17 common genes were screened, with 6 genes up-regulated and 11 genes down-regulated. The major functional annotations of differential genes were the generation of precursor metabolites and energy, Immune response, and Phosphorylation. They were concentrated on Focal adhesion, PI3K/Akt signaling pathway, and Human papillomavirus infection pathway. The expressions of VEGFA and THBS1 were down-regulated, while the expressions of ITGB1, MMP7, and SOX9 were up-regulated. The core genes VEGFA and THBS1 were highly expressed in Thyroid and Appendix, but lowly expressed in Testis. MMP7 was highly expressed in the Gall bladder and low in the Adrenal. SOX9 was highly expressed in Thyroid and lowly expressed in the bone marrow. ITGB1 was highly expressed in Fat and low in Pancreas. Bioinformatics technology can mine chip data and present new targets for diagnosing and treating DN, but further verification is needed.

Introduction

Diabetes mellitus is the most common metabolic systemic disease in clinical practice and has a high disability rate and mortality (1). Diabetic nephropathy (DN) is a serious chronic microvascular complication of diabetes, which is the main cause of ESRD. There are approximately 420 million diabetic patients worldwide, and 1 in 5 diabetic patients has DN. With the improvement of people’s quality of life, about 20-30% of patients with type 2 diabetes are accompanied by renal impairment (2,3). Under the continuous action and stimulation of high glucose levels, it will lead to the destruction of glomerular and renal tubular function. The formation of advanced glycation end products will deposit in the kidney, leading to the thickening of the glomerular basement membrane, mesangial cell proliferation, glomerulosclerosis, and renal interstitial fibrosis. Compared with high glucose stimulation, advanced glycation end products have a strong cytotoxic effect on islet β cells, which is an important reason for the continuous deterioration of diabetes (4-6). The main clinical manifestations of DN patients are hypertension, proteinuria, anemia, and renal failure, and can also cause complications such as myocardial infarction and neurogenic cystitis (7). However, the clinical symptoms of early DN patients are not obvious. In DN patients with metabolic syndrome, the metabolic disease will aggravate the degree of kidney damage, leading to a 5-year survival rate of less than 20% (8). Therefore, finding the potential pathogenic molecules and therapeutic targets of early DN is the focus of improving the quality of life of patients.

Gene chip, detecting differentially expressed genes between different samples, is an efficient and large-scale technology to obtain biological information (9). With the development and application of gene chip technology, many gene expression profile data have been generated. Mining valuable information from these data has become a hot topic in bioinformatics research. At present, many sequencing data are generated and stored in public databases, such as GEO and Cancer Genome Atlas (TCGA) (10). Founded in 2000, the GEO database is a gene expression database created and maintained by the National Center for Biotechnology Information (NCBI) in the United States. It is currently the largest public gene resource library, storing massive high-throughput gene expression data with comprehensive biological data and is easy to operate (11). The use of bioinformatics technology to collect, store, integrate, and analyze massive data provides more
possibilities for exploring deeper molecular mechanisms of disease occurrence and development (12). In addition, it is conducive to further exploring the molecular changes in disease progression and the functional connection of related signaling pathways. Bioinformatics technology was adopted. It aimed to provide new research ideas for elucidating the pathogenesis of DN and targeted therapy of the disease.

Materials and Methods

Data sources

“Diabetic kidney disease”, “Diabetic nephropathy”, “Expression profiling”, and “Homo sapiens” as key search terms, in the GEO database (http://www.ncbi.nlm.nih.gov/geo/), gene expression profile data were screened. The gene expression profiles of glomerular biopsy tissues including DN patients and normal healthy kidneys were selected as GSE30528, GSE96804, and GSE30122, respectively. In GSE30528, gene expression data of 9 DN and 13 normal renal tissues were included. Gene expression data of 41 DN and 20 normal renal tissues were contained in GSE96804 and gene expression data of 19 DN and 50 normal renal tissues in GSE30122.

Screening of differentially expressed genes

The gene expression chip data were downloaded from the GEO database, and the gene expression matrix was obtained by using the Oligo package in R language for background adjustment and data normalization. eBayes test was adopted, and the Limma package in R language was adopted to analyze the difference of merged datasets, selecting $| \log_{2} FC | > 0.5$. Due to a few different data sets combined, through removing batch effect, part of the genes of $| \log_{2} FC |$ value decreased. Therefore, $| \log_{2} FC | > 0.5$. $P<0.05$ as screening of the genes threshold. The Pheatmap package was adopted for the heatmap of the genes, and the ggplot2 package for the volcano map of the genes. The common genes in the three datasets were obtained by Venn diagram.

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Gene</th>
<th>Expression</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>51196</td>
<td>PLCE1</td>
<td>Down</td>
<td>Phospholipase C epsilon 1</td>
</tr>
<tr>
<td>53405</td>
<td>CLIC5</td>
<td>Down</td>
<td>Chloride intracellular channel 5</td>
</tr>
<tr>
<td>5800</td>
<td>PTPRO</td>
<td>Down</td>
<td>Protein tyrosine phosphatase receptor type O</td>
</tr>
<tr>
<td>259217</td>
<td>HSPA12A</td>
<td>Down</td>
<td>Heat shock protein family A (Hsp70) member 12A</td>
</tr>
<tr>
<td>199</td>
<td>AIF1</td>
<td>Down</td>
<td>Allograft inflammatory factor 1</td>
</tr>
<tr>
<td>2762</td>
<td>GMDS</td>
<td>Down</td>
<td>GDP-mannose 4,6-dehydratase</td>
</tr>
<tr>
<td>9037</td>
<td>SEMA5A</td>
<td>Down</td>
<td>Semaphorin 5A</td>
</tr>
<tr>
<td>2296</td>
<td>FOXC1</td>
<td>Down</td>
<td>Forkhead box C1</td>
</tr>
<tr>
<td>7057</td>
<td>THBS1</td>
<td>Down</td>
<td>Thrombospondin 1</td>
</tr>
<tr>
<td>7422</td>
<td>VEGFA</td>
<td>Down</td>
<td>Vascular endothelial growth factor A</td>
</tr>
<tr>
<td>2246</td>
<td>FGF1</td>
<td>Down</td>
<td>Fibroblast growth factor 1</td>
</tr>
<tr>
<td>3688</td>
<td>ITGB1</td>
<td>Up</td>
<td>Integrin subunit beta 1</td>
</tr>
<tr>
<td>1233</td>
<td>CCR4</td>
<td>Up</td>
<td>C-C motif chemokine receptor 4</td>
</tr>
<tr>
<td>2833</td>
<td>CXCR3</td>
<td>Up</td>
<td>C-X-C motif chemokine receptor 3</td>
</tr>
<tr>
<td>4316</td>
<td>MMP7</td>
<td>Up</td>
<td>Matrix metallopeptidase 7</td>
</tr>
<tr>
<td>6662</td>
<td>SOX9</td>
<td>Up</td>
<td>SRY-box transcription factor 9</td>
</tr>
<tr>
<td>9332</td>
<td>CD163</td>
<td>Up</td>
<td>CD163 molecule</td>
</tr>
</tbody>
</table>

Bioinformatics analysis of differentially expressed genes

The database for annotation, visualization, and integrated discovery (DAVID) software was applied for large-scale gene and protein biological function comprehensive information annotation. The most obviously enriched biological annotations can be found (13). Firstly, DAVID (http://www.pantherdb.org/) was adopted for functional annotation of the genes and bioinformatics analysis of pathway enrichment, and GO for functional annotation of the genes, through functional annotation, the functions of genes can be divided into molecular function, cellular component, and biological process, and KEGG for signaling pathway enrichment of the genes. The difference was considered statistically meaningful when $P<0.05$.

Core genes are functionally critical and have highly correlated relationships with other genes. For this purpose, String (http://www.string-db.org/) was adopted to draw and analyze the PPI network of the genes, subsequently, Cytoscape (version 3.9.1) software was to draw the PPI network, and the MCC algorithm in the CytoHubba plugin to screen the core genes. Based on the early screening core gene, from the NCBI Gene database (https://www.ncbi.nlm.nih.gov/gene/) download and analysis of core gene tissue expression data were performed.

Results

Screening results of differentially expressed genes

As illustrated in Figure 1A, the volcano map of the genes was drawn after the combination of GSE30528, GSE96804, and GSE30122 chip data. The number of up-regulated genes was higher, and 2,245 genes were obtained; Figure 1B suggests Venn diagram of the genes was drawn applying GSE30528, GSE96804, and GSE30122 chip data. The number of up-regulated genes was higher, and 2,245 genes were obtained; Figure 1B suggests Venn diagram of the genes was drawn applying GSE30528, GSE96804, and GSE30122 chip data. The number of up-regulated genes was higher, and 2,245 genes were obtained; Figure 1C and Table 1 suggest the expression heat map of 17 common genes in GSE30528, GSE96804, and GSE30122 chip data was drawn (each chip data was combined into one group). The expression of PLCE1, CLIC5, PTPRO, HSPA12A, AIF1, GMDS, SEMA5A, FOXC1, THBS1,
VEGFA, and FGF1 was clearly down-regulated, and the genes ITGB1, CCR4, CXCR3, MMP7, SOX9, and CD163 were clearly up-regulated in DN patients.

Functions and related pathways of differentially expressed genes

Figure 2 reveals the top 10 functions in the annotation of the genes were the generation of precursor metabolites and energy, Immune response, Phosphorylation, Organ development, Phosphorus metabolic process, Cell cortex, Germ cell migration, Phosphatidylinositol transporter activity, Nephrin development, Glycosaminoglycan binding, respectively.

Figure 3 indicates the top 10 signaling pathways of differentially expressed gene pathway enrichment were Focal adhesion, PI3K/Akt signaling pathway, Human papillomavirus infection, Complement and coagulation cascades, Ras signaling pathway, Pentose and glucuronate interconversions, Androgen and estrogen metabolism, Rap1 signaling pathway, Proteoglycans in cancer, ECM-receptor interaction, respectively.

PPI network construction of differentially expressed genes

Figure 4 suggests the PPI network of the genes, and it was found that VEGFA, AIF1, CD163, CXCR3, SEMA5A, FGF1, THBS1, ITGB1, MMP7, SOX9, TOXC1, and GMDS proteins had interactions. PLCE1 interacted with the PTPRO protein. Among these genes, CCR4 and HSPA12A had no obvious interaction with other proteins.

Tissue expression profiles of core genes

Figure 6 presents tissue expression profile data for core differentially expressed genes. The expression level of VEGFA was the highest in Thyroid, followed by Prostate, Lung, Endometrium, and Gall bladder. However, expression levels were relatively low in Testis, Ovary, Lymph nodes, Pancreas, and Brain (Figure 6A). The expression level of THBS1 was the highest in Appendix, followed by the Gall bladder, Endometrium, Urinary bladder, and Liver. However, expression levels were relatively low in Testis, Skin, Brain, Thyroid, and Salivary glands (Figure 6B). The expression of MMP7 was relatively high in the Gall bladder, Endometrium, and Urinary bladder. Nevertheless, it was vacant or poorly expressed in several tissues, such as the Adrenal, Thyroid, and Brain (Figure 6C). SOX9 expression was relatively high in Thyroid, Brain, Stomach, Salivary gland, and Prostate. However, it was not expressed in Bone marrow, Ovary, and Prostate, and showed low expression levels in Adrenal and other tissues (Figure 6D). The expression of ITGB1 is relatively high in Fat, Gall bladder, Endometrium, Placenta, and Urinary bladder. Nevertheless, low expression levels were observed in Pancreas, Bone marrow, Brain, Salivary gland, and Skin (Figure 6E).

Discussion

DN is a very common complication of diabetes, about 30-40% of diabetic patients will progress to DN, and the incidence is increasing year by year (14). DN is an inde-
activate multiple downstream signaling pathways through TLR4, including the nuclear transcription factor pathway, mitogen-activated protein kinase pathway, Janus kinase/signal transducer and activator of transcription pathway, phosphatidylinositol-3-kinase/protein kinase B pathway (21,22). Previous studies have shown that inflammatory response and abnormal expression of cytokines and chemokines play an important role in the pathogenesis of DN (23,24). Inflammatory response mediated by inflammatory factors can affect vascular endothelial dysfunction and glomerular endothelial cell damage, which is an important mechanism leading to diabetes (25). This is consistent with the potential mechanism for the involvement of immune-mediated inflammation in DN.

DN is the main cause of death in patients with type 2 diabetes mellitus, and it is also an important cause of inducing patients to progress to ESRD (26). VEGFA, THBS1, ITGB1, MMP7, and SOX9 were found to be the core genes in the pathogenesis of DN by the protein function expression network and MCC algorithm. VEGFA is a member of the PDGF/VEGF growth factor family, which is involved in the occurrence and development of a variety of tumor diseases (27). With the progress of DN, patients will have proteinuria, increased levels of serum creatinine and urea nitrogen, and progress to uremia, renal failure, etc. Silencing VEGFA can lead to the occurrence of diabetes, and induce massive proteinuria and advanced nodular glomerulosclerosis (28). In addition, VEGFA also plays a major role in maintaining normal kidney function, and the decrease of VEGFA content in glomeruli can lead to glomerular filtration barrier disorder, proteinuria, and renal failure (29). Patients treated with anti-VEGFA antibodies also show microangiopathies such as proteinuria, renal intimal lesions, and thrombosis (30). Exogenous recombinant THBS1 protein can inhibit liver lipid synthesis induced by a high-fat diet in mice, thus improving fatty liver (31). Moczulski et al. (32) proposed that THBS was clearly over-expressed in DN patients. Blocking THBS1-dependent TGF-β activation can selectively target excessive TGF-β activity and improve renal tubulointerstitial damage in DN models (33).

Inflammatory treatment can effectively protect the normal operation of renal function, thereby delaying the progression of DN. PI3K/Akt/mTOR pathway is important in a variety of inflammatory diseases, including chronic obstructive pulmonary disease, rheumatoid arthritis, and inflammatory bowel disease (34). Silencing ITGB1 can inhibit the activation of the FAK/PI3K/Akt pathway and inhibit the activity of retinal microvascular endothelial cells and retinal Müller cells in diabetic retinopathy (35). However, the binding of the ITGB1 promoter to the transcription factor GATA can lead to the expression of circITGB1 and enhance the inflammatory response induced by renal hypoxia/reperfusion through the miR-328-3p/PIM1 axis (36). Inhibition of ITGB1 expression can alleviate hypoxia/reoxygenation-induced renal cell apoptosis and inflammatory response (37). The association between ITGB1 and DN is unknown. MMP7 can promote renal fibrosis in DN patients, while the inactivation of MMP7 attenuates the mesenchymal transition of renal tubular epithelial cells induced by high glucose (38). Some studies have suggested that the urinary MMP7 content of DN patients is closely correlated with the mortality of patients (39).
SOX9 is expressed in renal tubules, and the continuous expression of SOX9 is closely associated with renal fibrosis (40). SOX9 overexpression can lead to ectopic expression of aggrecan and type II collagen in mesangial cells, and then participate in the progression of DN (41-46). In addition, VEGFA and THBS1 were highly expressed in the Thyroid and Appendix, respectively, and lowly expressed in the Testis. MMP7 was highly expressed in the Gall bladder and low in the Adrenal. SOX9 was highly expressed in Thyroid and lowly expressed in the bone marrow. ITGB1 was highly expressed in Fat and poorly expressed in Pancreas, suggesting that VEGFA, THBS1, ITGB1, MMP7, and SOX9 may involve in the occurrence and development of DN, but it needs more experiments to verify.

Conclusion
17 genes in DN and normal renal tissues were screened, and they were mainly involved in the inflammatory response. VEGFA, THBS1, ITGB1, MMP7, and SOX9 are the core genes in the pathogenesis of DN, and their abnormal expression may be involved in the progression of DN. Only the public datasets in the GEO database were used to screen the potential pathogenic molecules of DN, and clinical data or experimental design are needed to verify the mechanism of core genes and DN progression. In conclusion, these results provide a bioinformatics basis for understanding the pathogenesis of DN, achieving an early diagnosis of the disease, and developing targeted therapeutic drugs.

Fundings
The research is supported by: Key R&D projects in Heilongjiang Province, (No. 2022ZX06C16).

References


56