Hub Genes Associated with the Diagnosis of Diabetic Retinopathy

Yanhui Tang1,2,*
Qi Tang1,2,*
Haicheng Wei1,2,*
Pinzhang Hu1,2
Donghua Zou3
Rixiong Liang1,2
Yu Ling1,2

1Department of Ophthalmology, The Fifth Affiliated Hospital of Guangxi Medical University, Nanning, Guangxi, 530022, People’s Republic of China;
2Department of Ophthalmology, The First People’s Hospital of Nanning, Nanning, Guangxi, 530022, People’s Republic of China;
3Department of Neurology, The Fifth Affiliated Hospital of Guangxi Medical University, Nanning, Guangxi, 530022, People’s Republic of China

*These authors contributed equally to this work

Purpose: This study aimed to identify genes that may be effective in diagnosing or treating diabetic retinopathy (DR), the most common complication of diabetes mellitus (DM).

Methods: Differentially expressed genes (DEGs) were identified between DR and DM in GSE146615 dataset. DEGs that were consistently up- or down-regulated under both standard glucose and high glucose conditions were identified as common genes and used to generate a protein–protein interaction network and modules. The module genes were assessed for the area under the receiver operating characteristic curve (AUC), leading to the identification of hub genes. Differentially methylated probes in GSE76169 were also compared with common DEGs to identify specific methylation markers of DR. Enrichment analysis was used to explore the biological characteristics. The Short Time-series Expression Miner algorithm was used to identify genes that were progressively dysregulated in the sequence: healthy controls < DM < DR.

Results: A total of 1917 common genes were identified for seven modules. The eight genes with AUC > 0.8 under high glucose and standard glucose conditions were considered as hub genes. The module genes were significantly enriched during vascular smooth muscle cell development and regulation of oxygen metabolism, while 92 methylation markers were involved in the similar terms. Among the progressively dysregulated genes, three intersection genes under both standard glucose and high glucose conditions were found to be module genes and were considered as key genes.

Conclusion: We identified eight potential DR-specific diagnostic and therapeutic genes, whose abnormal expression can cause oxidative stress, thus favoring the course of the disease.

Keywords: diabetic retinopathy, oxidative stress, glucose, cluster analysis, hub genes

Introduction

Diabetes mellitus (DM) affects an increasing number of people worldwide.1 The global incidence of DM is estimated to be 415 million adults and is expected to be nearly 700 million by 2045.2,3 The course of DM is closely related to the development and progression of diabetic retinopathy (DR),4 which is one of the most common complications of DM and which can reduce visual acuity and even lead to blindness.5,6 Almost all patients with type 1 diabetes mellitus (T1DM) and more than 60% with type 2 diabetes mellitus (T2DM) have some degree of retinopathy after 20 years with DM.7 The estimated global prevalence of DR is 382 million,8 imposing a heavy financial burden on healthcare systems.9,10

The retina is the most active metabolic tissue and consumes a large amount of energy and oxygen.11 Inadequate oxygen and nutrient supply to the retina may impair
the oxygen balance in cells, resulting in a hypoxic retinal environment.\textsuperscript{12} During the development of DM, high blood glucose reduces the oxygen levels in the retina,\textsuperscript{13} while chronic hyperglycemia alters the function of vascular smooth muscle cells, thus promoting DR.\textsuperscript{14} A typical manifestation of DR is gradual change in retinal microvasculature, which in turn leads to retinal ischemia, neovascularization, changes in retinal permeability, and macular edema.\textsuperscript{15} Despite the existing knowledge about the pathology and manifestation of DR, no prophylactic treatment has been reported for the early stages of the disease.\textsuperscript{16}

DR can be clinically divided into non-proliferative diabetic retinopathy (NPDR), which is characterized by microvascular damage, and proliferative diabetic retinopathy (PDR), where new blood vessels are formed.\textsuperscript{17} NPDR can also develop into PDR, resulting in irreversible vision loss and complete blindness.\textsuperscript{18} Current treatments for DR include laser surgery and anti-vascular endothelial growth factor (VEGF) therapy, which can control the growth of abnormal blood vessels but cannot prevent blindness.\textsuperscript{19} Gene therapy can also be used as a restorative or neuroprotective approach depending on when it is applied during the disease and what genes it targets.\textsuperscript{20} DNA methylation in DM patients has also been associated with the pathogenesis of DR, suggesting that methylated sites can serve as predictive or diagnostic biomarkers of DR.\textsuperscript{21}

Here, we used the GSE146615 dataset to identify potential diagnostic markers and therapeutic targets in DR, as well as to begin to clarify molecular dysregulation in the disease. In addition, we used the GSE76169 dataset to explore potential associations between methylation mechanisms and DR. In GSE76169 dataset, all T1DM cases received conventional treatment in the Diabetes Control and Complications Trial (DCCT)\textsuperscript{22} and developed retinopathy at the 10th year after the end of the EDIC. In contrast, controls were in the DCCT intensive treatment group and did not develop retinopathy during the EDIC study. Identifying early features of DR is particularly important for inhibiting pathology and preventing visual impairment. Therefore, we expect that understanding the underlying pathogenic mechanisms of DR will contribute to the development of novel therapeutic approaches.

Materials and Methods

Data Processing
The DR gene data were collected from the Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov). The GSE146615 dataset included gene expression profiles of lymphoblastoid cell lines derived from healthy controls (NoD) and matched T1DM individuals with retinopathy (DwC) and without retinopathy (DwO). The cells were cultured under standard glucose (SG) (21 NoD, 26 DwC, and 25 DwO) and high glucose (HG) conditions (21 NoD, 26 DwC, and 25 DwO), and all samples were analyzed on Illumina HT12v4 microarrays (ThermoFisher).

The raw data were background-subtracted and normalized using the lumiR package.\textsuperscript{23} The GSE76169 dataset included the whole-blood DNA methylation profiles from 32 DR cases and 31 controls collected at the beginning of the Epidemiology of Diabetes Interventions and Complications (EDIC) study.\textsuperscript{24} DNA methylation samples were measured on the Illumina HumanMethylation450 BeadChip array, and the Bioconductor package ChAMP was used to generate the normalized beta values.\textsuperscript{25} All data in the present study were obtained from an open-access database and no ethical approval was required.

Differential Analysis
The limma package inR\textsuperscript{26} was used for the differential analysis between DR and DM patients or between DM patients and controls in the GSE146615 dataset under HG or SG conditions. Genes with \( P \) value < 0.05 of \( t \) test were considered statistically significant and were indicated as differentially expressed genes (DEGs). In contrast, common genes were defined as genes whose expression in DR differed in the same direction relative to DM under both HG and SG conditions. The limma algorithm was also applied to the GSE76169 dataset to obtain the differentially methylated CpG sites with \( P < 0.05 \).

Network Construction
The Search Tool for Retrieval of Interacting Genes/Proteins (STRING; http://string-db.org) is an online database for analyzing functional networks between two or more proteins. Here, the protein–protein interaction (PPI) network was constructed by analyzing common genes with STRING based on the criterion of a combined score >500. The network was displayed using Cytoscape software. Significant modules in the PPI network were identified by the molecular complex detection (MCODE) plugin in Cytoscape\textsuperscript{27} with degree cutoff = 2, node score cutoff = 0.2, K-Core = 5, and max. depth = 100.
Enrichment Analysis

Annotation of signaling pathways and functions for module genes were performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID, http://david.ncifcrf.gov). Genes were also analyzed for enrichment of Gene Ontology (GO) processes and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways pathways using the clusterProfiler package in R. The same package was also used within a Java environment to obtain the background set for gene set enrichment analysis (GSEA) between DR and DM. The fast preranked gene set enrichment analysis (fgsea) package in R was used to further analyze the GSEA results. Differences with $P$ value $< 0.05$ were considered statistically significant. Gene set variation analysis (GSVA) was carried out using the GSVA package to explore the activation or inhibition of biological processes and KEGG pathways in DR relative to DM.

Short Time-Series Expression Miner Analysis

Short Time-series Expression Miner (STEM) software was used to cluster module genes of the GSE146615 dataset in healthy, DwoC, and DwC samples under SG and HG conditions. A $P$ value $< 0.05$ was considered as statistically significant clustering. The expression of the significantly clustered genes was progressively up- or down-regulated in the sequence: healthy < DwoC < DwC. The area under the receiver operating characteristic curve (AUC) was also calculated to evaluate the diagnostic value of the hub genes using the pROC package.

Single-Sample GSEA

The marker gene sets defining different immune cell types were obtained from a previous study. Differences in infiltration levels by immune cells between DR and DM was calculated using the single-sample gene set enrichment analysis (ssGSEA) function in the GSVA package and was expressed as a ssGSEA score. A $P$ value $< 0.05$ was considered statistically significant.

Results

Identification of DEGs Between DR and DM

In order to identify specific genes associated with DR, differential analysis of gene expression between DR and DM in GSE146615 was first performed (Figure 1), resulting in 3155 DEGs under HG conditions (Figure 2A, Table S1) and 3045 DEGs under SG conditions (Figure 2B, Table S2). The intersection of DEGs were attributed to the progression of DR and indicated that 1053 common genes were upregulated and 864 were downregulated under the two glucose conditions (Figure 2C).

PPI Network

The protein interactions among the identified common genes were predicted using the STRING database. In particular, the

Figure 1 Study flowchart. The GSE146615 dataset was used to identify potential diagnostic biomarkers and therapeutic targets of diabetic retinopathy (DR), as well as molecular mechanisms of the disease. Alterations in methylation of DR-related genes were also screened to identify methylated markers.

Abbreviations: HG, high glucose; SG, standard glucose; PPI, protein–protein interaction; STEM, Short Time-series Expression Miner; MCODE, molecular complex detection; GSVA, gene set variation analysis; AUC, area under the receiver operating characteristic curve.
1917 common genes were searched in the STRING (http://string-db.org) and screening for interaction scores >500 yielded a network of 6775 reciprocal pairs of PPIs. In addition, Cytoscape software was used to identify the interaction relationship of the PPI network genes, and the MCODE plugin recognized 279 module genes (Figure 3A). Further calculation of their AUC values under HG and SG conditions indicated that the following eight genes had AUC > 0.8: FOXO1, ALDH9A1, TSC1, AHI1, CMKLRI, EXOSC7, INTS6, and MDFIC. These eight genes were identified as hub genes (Figure 3B). The differential expression of these genes between DR and DM was similar under SG conditions (Figure 3C) and HG conditions (Figure S1).

The infiltration by immune cells in DR and DM was also quantified, revealing differences between the two conditions. TH2 and B cells were significantly downregulated in DR compared to DM under both SG (Figure 3D) and HG (Figure S2) conditions. Correlation analysis also showed that infiltration by TH2 cells correlated positively with AHI1 expression and negatively with TSC1 expression under both glucose conditions (Figure 3E and Figure S3). The infiltration by B cells correlated positively with MDFIC expression and negatively with FOXO1 expression.

**Enrichment Analysis of Module Genes**

The module genes were studied by enrichment analysis to identify the molecular mechanisms of DR. There were 414 terms in the DAVID analysis results, including 5’-adenosine monophosphate-activated protein kinase (AMPK) signaling pathway, cell cycle, and tumor necrosis factor (TNF) signaling pathway (Table S3). In the enrichment results, the activation or inhibition of biological processes and KEGG terms in DR relative to DM was examined under SG conditions using the GSVA package. Specifically, vascular smooth muscle cell development, regulation of retinoic acid biosynthetic process, and regulation of oxygen metabolism were activated in DR and significantly enriched in module genes (Figure 4A). In contrast, the
activation of c-Jun N-terminal (JUN) kinase, mitogen-activated protein kinase (MAPK), and chronic inflammatory response were inhibited in DR. The KEGG pathways of valine, leucine, and isoleucine degradation, AMPK signaling, and β-alanine metabolism were also activated in DR and significantly enriched in module genes (Figure 4B), whereas the cell cycle and signaling pathways of Toll-like receptor (TLR) and TNF were inhibited. Furthermore, GSEA indicated the activation of histidine metabolism, maturity-onset diabetes of the young, and neutrophil extracellular trap formation under SG (Figure 4C) and HG (Figure 4D) conditions in DR. GSEA also indicated inhibition of the T-cell receptor, TLR, and NFκB signaling pathways in DR.

**Methylation of Common Genes**

The methylation profiles of DR cases and controls in the GSE76169 dataset were studied by differential analysis to
determine if the expression of common genes might be regulated by methylation. Genes were considered to be methylation markers regulated by methylation when their expression differed (under SG conditions) in the opposite direction to methylation differed in DR patients compared controls. A total of 2992 differentially methylated probes (DMPs) were identified (Figure 5A) and after comparing them with common genes, we found 92 methylation markers that may be involved in the progression of DR (Figure 5B).

Enrichment analysis revealed that these methylation markers were significantly involved in biological processes, such as “vascular smooth muscle cell development”, “reactive oxygen species biosynthetic process” (Figure 5C), In KEGG results, we identified significantly enriched “Notch signaling pathway”, “Glycine”, “serine and threonine metabolism” (Figure 5D).

**Potential Genes Involved in DR Progression**

Through STEM analysis, we identified the genes whose persistent expression changes during the progression from healthy controls to DM patients and then to DR patients, under both HG and SG conditions. A total of 18 and 17 common genes showed progressively up-or down-regulated expression under SG and HG conditions, respectively, and the groups of genes fell into two clusters (Figure 6A and B). Among the two groups of progressively dysregulated genes, three intersection genes were found to be module genes of the PPI network (CXCL10, HCK, and HLA-DRB1), and were therefore considered as key genes. The expression of key genes was progressively downregulated from healthy controls to DM patients and then to DR patients, under both SG (Figure 6C) and HG (Figure 6D) conditions.

**Discussion**

The identification of effective biomarkers for DR is particularly important, as such biomarkers may contribute to the early diagnosis and treatment of the disease. Although biomarkers with therapeutic potential are commonly used in medicine and especially in oncology, none has been reported for DR. In the present study, we investigated potential diagnostic and therapeutic

---

**Figure 4** Module genes involved in biological processes and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. (A and B) Numbers of module genes in diabetic retinopathy compared to healthy controls involved in activation or inhibition of biological processes and KEGG pathways, as determined by Gsva. Longer columns indicate the involvement of more genes under the given term. (C and D) Module genes involved in activation or inhibition of KEGG pathways in diabetic retinopathy compared to healthy controls under (C) standard glucose or (D) high glucose conditions, as determined by GSEA. Differences with $p < 0.05$ were considered statistically significant.

**Abbreviations:** Gsva, gene set variation analysis; GSEA, gene set enrichment analysis; BP, biological processes; NES, normalized enrichment score.
biomarkers of DR by comparing gene expression changes in DR and DM cell lines under SG and HG conditions.

Screening of genes with different expression in the same direction under SG and HG conditions revealed that several genes may play a key role in DR diagnosis. In contrast to DM, FOXO1, ALDH9A1, and TSC1 are upregulated in DR, thus promoting the disease. Specifically, the upregulated expression of FOXO1 causes IL-1β-induced autostimulation, and its upregulation in endothelial cells significantly contributes to the development of DR. In addition, FOXO1 acts downstream of insulin-dependent signaling pathways, which are dysregulated in DM. TSC1 is another important gene that contributes to the development of diabetic complications, including DR, by targeting the mechanistic target of rapamycin (mTOR) pathway. Among the genes downregulated in DR, AH1I encodes a cilia-localized protein, and its absence promotes retinal degeneration. CMKLR1 is activated by the Wnt/β-catenin pathway, which is involved in extracellular matrix accumulation and renal fibrosis in DR. However, the pathogenic role of CMKLR1 in retinal microvascular function has not been evaluated in vivo or in vitro. EXOSC7 has been reported to be differentially expressed between T1DM and healthy controls, but no study has been found showing its altered expression in DR. INTS6 is a glucocorticoid response gene, and glucocorticoids have gluconeogenesis effects on the liver. Levels of INTS6 are significantly upregulated after intermittent fasting in patients with metabolic syndrome. The glucocorticoid receptor binds to MDFIC, and directly inhibits its expression. And MDFIC was significantly downregulated in retinopathy compared with the control. Although it is clear that the up- and down-regulation of hub genes plays a significant role in the DR process, their specific roles in the disease remain to be elucidated. In addition, the hub genes may

---

**Figure 5** Identification of methylation markers for common genes. (A) Differentially methylated positions between diabetic retinopathy and diabetes mellitus. Among them, 71.19% were hypermethylated and 28.81% were hypomethylated. Different colors represent different methylation sites. Labels indicate the 10 sites with the highest methylation levels. (B) Methylation and expression levels of methylation markers. The labels indicate the important genes. (C) Biological functions and (D) KEGG pathways significantly enriched by methylation markers. Thicker lines represent greater correlation between the terms and methylated markers.

**Abbreviations:** DMPs, differentially methylated positions; KEGG, Kyoto Encyclopedia of Genes and Genomes; corr, Pearson correlation coefficient.
Figure 6 Identification of progressively dysregulated genes during the development of diabetic retinopathy. (A and B) Progressively dysregulated genes under (A) standard glucose (SG) and (B) high glucose (HG) conditions were identified by STEM analysis in the trend: healthy controls < diabetes mellitus < diabetic retinopathy. The genes were then clustered into a progressively upregulated cluster (Cluster 1) and a progressively downregulated cluster (Cluster 2). (C and D) Expression of CXCL10, HCK, and HLA-DRB1 genes among healthy controls, diabetic retinopathy, and diabetes mellitus under (C) SG and (D) HG conditions. Abbreviations: STEM, Short Time-series Expression Miner; NoD, healthy controls; DwC, type 1 diabetes individuals with retinopathy; DwoC, type 1 diabetes individuals without retinopathy.
affect the immune response of DR, according to our correlation results between immune cell infiltration and gene expression.

Here, we first performed enrichment analysis of module genes within the PPI network and identified potential biological functions and signaling pathways that may be involved in the pathophysiology of DR. Oxidative metabolic processes are strongly activated in DR cells, as hypoxia in the retina can induce neuronal degeneration. Alterations in the excitability of vascular smooth muscle cells play a key role in the development and progression of vascular complications in DM. Animal studies have shown that hyperglycemia can cause vascular dysfunction in DR, while inflammation has also been indicated as a key factor in the development of DR. However, here, we found that chronic inflammation is suppressed during DR, suggesting that the inflammatory response of in vitro cultured DR cells may not play a major pathological role in the disease development.

Enrichment analysis of KEGG terms showed that the AMPK signaling pathway is activated in DR. This pathway mediates the stress response, promotes autophagy, and stabilizes hypoxia-inducible factor 1α (HIF-1α), thus enhancing VEGF expression. VEGF hypersecretion then leads to irregular retinal vasculature, resulting in edema, hemorrhage, and eventually retinal detachment. However, we found that inflammatory signaling pathways, such as those involving TLR and TNF, are suppressed in DR, which is inconsistent with previous studies that suggested that HG levels enhance the expression of TLR, NFκB, and TNF-α, thereby exacerbating DR. Our results suggest that DR cells cultured in vitro may be more prone to activate oxidative stress-related signals, while suppressing inflammation-related responses. This may be related to the reduced immune cell content in DwC compared to DwoC of GSE146615 dataset.

STEM analysis showed that the expression of CXCL10, a known biomarker for DR, as well as expression of HCK and HLA-DRB1 was progressively down-regulated from healthy controls to DM patients and then to DR patients. This suggests that the three genes may be involved in the progression of DR in patients with DM. These results are inconsistent with recent studies, where CXCL10 expression was upregulated from DM to DR. This may be related to the suppression of inflammatory response in DR in our results. Platelet-derived vesicles seem to regulate hyperglycemia-induced retinal endothelial damage by releasing CXCL10, which can then activate the TLR4 pathway. HCK belongs to the Src family kinase involved in regulating endothelial cell proliferation, migration, and angiogenesis, and it is differentially expressed between DM and normal subjects, as well as between healthy controls and patients with diabetic peripheral neuropathy. However, its role in DR progression has not yet been elucidated. HLA-DRB1, for its part, is a vision-related gene that is differentially expressed in proliferative DR lesions, while its polymorphisms are known risk factors for DM and its complications. Based on these earlier findings and our knowledge of DR pathophysiology, we suggest that CXCL10, HCK, and HLA-DRB1 play an important role in the development and progression of the disease.

Differential analysis of the methylation profiles of DR cases and controls in the GSE76169 dataset identified altered DNA methylation of DR-related genes. Consistent with our results, transcription of genes related to redox balance is known to be regulated by cytosine methylation status in DNA. This methylation is involved in retinal development and responds dynamically to environmental changes. Indeed, the restoration of proper DNA methylation can restore mitochondrial homeostasis, maintain cellular redox balance, and inhibit or delay DR progression. In the KEGG pathways of methylation markers enriched, the balance of Notch signaling controls the formation and degeneration of retinal vessels. Hyperglycemia induced activation of serine/threonine kinase family proteins is one of the factors involved in microvascular alterations in DM that can lead to DR.

Our results should be interpreted with caution. The small sample in our study means that our findings should be verified and extended in larger studies. Moreover, our findings were not validated through molecular or clinical experiments, which will be performed in subsequent studies to confirm the efficacy of the identified genes as novel diagnostic biomarkers or therapeutic targets in DR. In addition, although T1DM and T2DM overlap in some clinical and pathological features, their molecular mechanisms and genetic basis are different. We only selected T1DM individuals to do the analysis, whether these results are applicable in T2DM needs further study.

Conclusion
Our study on differentially expressed genes between DR and DM showed that oxidative stress may be an important
pathological mechanism in DR cells cultured in vitro. A total of eight genes were identified that could be used as valuable biomarkers for DR, while CXCL10, HCK, and HLA-DRB1 could serve as biomarkers of DR severity.

Abbreviations
AMPK, monophosphate-activated protein kinase; AUC, area under the receiver operating characteristic curve; DAVID, Database for Annotation, Visualization and Integrated Discovery; DCCT, Diabetes Control and Complications Trial; DEGs, differentially expressed genes; DM, diabetes mellitus; DMPs, differentially methylated probes; DR, diabetic retinopathy; DwC, type 1 diabetes individuals with retinopathy; DwoC, type 1 diabetes individuals without retinopathy; EDIC, Epidemiology of Diabetes Interventions and Complications; fgssea, fast pre-ranked gene set enrichment analysis; GO, Gene Ontology; GSEA, gene set enrichment analysis; GSVA, Gene set variation analysis; HG, high glucose; HIF-1α, hypoxia-inducible factor 1α; JUN, c-Jun N-terminal kinase; KEGG, Kyoto Encyclopedia of Genes and Genomes; MAPK, mitogen-activated protein kinase; MCODE, molecular complex detection; NoD, healthy controls; NPDR, non-proliferative DR; PDR, proliferative DR; PPI, protein–protein interaction; SG, standard glucose; ssGSEA, single-sample gene set enrichment analysis; STEM, Short Time-series Expression Miner; STRING, the Search Tool for Retrieval of Interacting Genes/Proteins; TLR, Toll-like receptor; TNF, tumor necrosis factor; VEGF, vascular endothelial growth factor.

Data Sharing Statement
Data were downloaded from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/).

Funding
This study was supported by the High-Level Medical Expert Training Program of Guangxi “139” Plan Funding (G201903049).

Disclosure
The authors report no conflicts of interest related to this work.

References


42. Louie CM, Caridi G, Lopes VS, et al. AH1 is required for photo-receptor outer segment development and is a modifier for retinal degeneration in nephronophthisis. *Nat Genet.* 2010;42(2):175–180. doi:10.1038/ng.519


