

Diet-induced Oxidative Stress as a Driver of Diabetes Progression and Diabetic Complications: Insights from Single-Cell and Transcriptomic Analyses

Abstract: Diabetic nephropathy (DN) and diabetic retinopathy (DR) are major microvascular complications of diabetes that exhibit striking pathological similarities, yet their shared molecular basis remains unclear. Given the central role of oxidative stress in diabetic vascular injury, this study aimed to identify key regulatory factors contributing to endothelial dysfunction and to elucidate the shared molecular network linking renal and retinal damage. Human transcriptomic datasets from DN and DR were integrated to identify differentially expressed genes and construct disease-associated co-expression modules. These findings were further combined with single-cell RNA sequencing data derived from a high-fat high-fructose-induced oxidative stress model to characterize endothelial cell-specific transcriptional signatures. Intersection analysis revealed a conserved gene subset shared across datasets. Functional enrichment demonstrated that these genes were primarily involved in immune activation, extracellular matrix remodeling, complement and coagulation cascades, and vascular structural regulation. The LASSO analysis four hub genes (*NOTCH1*, *CELF2*, *EMP3*, and *VGLL4*) as central candidates associated with mediated endothelial dysfunction. A diagnostic model constructed based on these four genes demonstrated stable and reproducible performance across independent datasets.

In conclusion, my research results indicate that diabetic nephropathy (DN) and diabetic retinopathy (DR) are organ-specific manifestations of a common oxidative stress-driven endothelial injury program, providing mechanistic insights and potential molecular targets for the dual organ treatment strategy of diabetes.

Key words: Oxidative stress; Diabetic nephropathy; Diabetic retinopathy; Endothelial dysfunction

1 Introduction

Diabetes mellitus represents one of the most pressing chronic diseases worldwide, for which no curative therapy currently exists. Over recent decades, its prevalence has increased dramatically across both developed and developing nations, creating a substantial global health challenge^[1]. Projections from the World Health Organization indicate that by 2030, diabetes will rank as the seventh leading cause of death worldwide^[2]. The International Diabetes Federation (IDF) estimated the global population with diabetes mellitus (DM) to be 463 million in 2019 and 700 million in 2045^[3]. This escalating burden carries profound socioeconomic consequences, including elevated healthcare expenditures, reduced workforce productivity and so on. Therefore, the pressing demand for more effective treatments has spurred a surge in research aimed at unraveling the molecular underpinnings of diabetes and its devastating complications.

The majority of diabetes-related mortality stems from its vascular complications, with diabetic nephropathy (DN) and diabetic retinopathy (DR) representing the most prevalent and severe manifestations^[3,4]. In diabetic patients, these two complications often occur concomitantly and share marked pathological similarities, such as basement membrane thickening, extracellular matrix deposition, and progressive microvascular dysfunction^[5,6]. Such consistency implies the existence of common mechanistic pathways that transcend tissue-specific regulators. The pathogenesis of diabetic microangiopathy is multifactorial, encompassing metabolic disorders, hemodynamic changes, oxidative stress, and chronic low-grade inflammation^[7]. Of these factors, microvascular endothelial dysfunction represents a critical early event in both diabetic nephropathy and diabetic retinopathy, typified by compromised nitric oxide bioavailability, elevated vascular permeability, and a prothrombotic phenotypic transition^[8]. Therefore, understanding the molecular basis of endothelial dysfunction may provide a unified framework for elucidating disease mechanisms.

Despite the established role of endothelial dysfunction as a central pathogenic event, the specific molecular regulators driving this process remain poorly defined. This study identifies reactive oxygen species (ROS) as a critical mechanistic nexus. ROS-induced oxidative stress elicits lipid peroxidation, genotoxicity, and pro-inflammatory signaling, culminating in glomerular endothelial injury and retinal vascular leakage^[9-11]. These parallel pathologies implicate ROS-mediated endothelial damage as a shared inciting event in diabetic microangiopathy. Therefore, this study proposes a hypothesis: by characterizing the key regulatory genes in the reactive

oxygen network, potential biomarkers that can exert intervention and therapeutic effects on both DN and DR will be discovered.

This study employed a comprehensive multi-omics approach, integrating single-cell RNA sequencing data generated from a diet-induced oxidative stress rat model with the overall transcriptome data from human samples of diabetic nephropathy and diabetic retinopathy. The animal model allows for the dissection of cellular heterogeneity and tissue-specific adaptive responses under metabolic stress, whereas human clinical datasets furnish disease-associated molecular signatures. Through cross-validation between bulk and single-cell transcriptomic platforms, to reveal the common molecular mechanisms driven by oxidative stress behind renal and retinal microvascular complications. This cross-complication and cross-technology analytical strategy provides a unique insight into diabetic microangiopathy, with the potential to uncover previously unrecognized pathways linking DN and DR, as well as to identify novel biomarkers and therapeutic targets directed against oxidative stress and endothelial dysfunction.

2 Materials and Methods

2.1 Data source

For the microarray transcriptome analysis, kidney tissue samples from diabetic nephropathy (DN) patients and healthy controls obtained from the NCBI GEO GSE30122 dataset^[1].

For the analysis of RNA sequencing (RNA-seq) data, this project selected 10 samples of patients with diabetes but without complications, 10 samples of patients with diabetes and showing retinopathy (DR), and 5 normal control samples from the NCBI GEO GSE221521^[2] dataset for download and analysis.

For single-cell RNA sequencing (scRNA-seq) data analysis, different tissues from high-fat high-fructose (HFHF)-fed aged rats and control aged rats from the NCBI GEO GSE201900 dataset^[4] was analyzed.

2.2 Processing SRA data to obtain gene and cell matrices

2.2.1 Quality control of FASTQ files

Quality control of sequencing data was conducted using FastQC (v0.11.9), generating QC reports that evaluated multiple metrics including per-base mean sequence quality, per-sequence quality scores, per-base sequence content, average sequence GC content, per-base N content, sequence length distribution, relative sequence duplication levels, overrepresented sequences and adapter content.

Aggregated reports from MultiQC showed that per-base sequence content and sequence duplication levels failed to meet conventional quality criteria, indicating a requirement for further data filtering.

2.2.2 Data filtering and trimming

Based on the quality assessment results, low-quality reads and adapter sequences were removed using Trimmomatic (v0.39) with the following parameters: LEADING:3, TRAILING:3, SLIDINGWINDOW:4:15, HEADCROP:12, and MINLEN:36. This step ensured the removal of low-quality bases, adapters, and short fragments, providing a clean dataset for subsequent analysis.

2.2.3 Alignment to the reference genome (Alignment of fragmented sequences)

Clean paired-end reads were aligned to the reference genome (hg38) using HISAT2 (v2.2.1) with the --dta parameter to optimize downstream transcript assembly.

The resulting Sequence Alignment Map (SAM) files were converted to binary BAM format using SAMtools (v1.18) and sorted for further processing.

2.2.4 Quantification of read counts

The aligned sequencing reads were counted using FeatureCounts (v2.0.6). Raw expression counts for all genes in each sample were obtained, resulting in the raw expression matrix.

2.2.5 Cell Ranger pipeline analysis

The filtered feature-barcode matrix output by the Cell Ranger (v 10.0.0) pipeline was directly loaded and processed into a Seurat object. All samples are represented by the barcode sequence for individual cells, and the UMI count results indicate the mRNA expression levels in the cells.

2.3 Acquisition of expression matrix from microarray

The raw data of the DN microarray dataset were subjected to preprocessing to obtain a normalized gene expression matrix. Background correction and quantile normalization were performed, and probe-level intensities were summarized into gene-level expression values using the robust multi-array average (RMA) algorithm. Probes were annotated to corresponding gene symbols according to the microarray platform.

2.4 Differential expression analysis between groups

Differential expression analysis was performed separately for the DN (microarray) and DR (RNA-seq) datasets. For the DN dataset, analysis was conducted using the limma R package(v 3.64.3) with moderated t-tests. For the DR dataset, the DESeq2 package was applied to the TMM-normalized count data. In both analyses, genes meeting the criteria of an adjusted p-value < 1 and an absolute log₂ fold change > 0.05 were defined as significantly differentially expressed genes (DEGs).

2.5 Weighted gene co-expression network analysis (WGCNA)

Weighted gene co-expression network analysis was performed on the DN microarray dataset and DR RNA-seq dataset, respectively. The optimal soft threshold (power value) was calculated and determined according to the scale-free network criterion to construct the gene co-expression network and divide the genes into different co-expression modules. The correlation between each module and the phenotypic traits was calculated, and the correlation heatmap and scatter plot were drawn to screen the key modules with significant correlation with the phenotype of diabetic microangiopathy.

2.6 Seurat object creation and cell filtering

scRNA-seq data from samples were processed using the Seurat package (v5.3.1) in R. A combined Seurat object was created by merging the raw gene expression matrices from all samples. Stringent quality control was applied to remove low-quality cells and genes. Cells were filtered based on the following criteria: retained cells required a detected gene count between 200 and 2,000, and a mitochondrial gene percentage below 10%. Genes not expressed in at least 0.5% of the total cell population were excluded.

2.7 Dimensionality reduction and clustering

Principal component analysis (PCA) was performed on the normalized and scaled data using the RunPCA function in the Seurat package. The top 2,000 highly variable genes were selected as input for PCA. An elbow plot was then generated to visualize the proportion of variance explained by each principal component. Based on the inflection point observed in the elbow plot, the first 13 principal components (PCs) were retained for subsequent analysis. Cell-cell similarities were calculated using the FindNeighbors method based on Euclidean distance in the PCA space. Cell clusters were identified using the Louvain algorithm via the FindClusters function at a resolution of 0.7. Finally, the resulting clusters were visualized in two dimensions using the Uniform Manifold Approximation and Projection (UMAP) method.

2.8 Identification of DEGs across clusters

To identify marker genes for each cell cluster, differential expression analysis was performed using the FindAllMarkers function in Seurat. Genes were considered significant markers if they met the following thresholds: expressed in a minimum of 25% of cells within the cluster, displayed an average log₂ fold change greater than 0.5 compared to all other clusters, and exhibited an adjusted p-value less than 0.05. The top three marker genes ranked by average log₂ fold change were selected for each cluster for visualization and annotation purposes.

2.9 Cell type annotation and sub-clustering

The identification of cell types was carried out using the method of manual annotation. Eight major cell types were defined by using classical marker genes. Given its significant role in diabetic microangiopathy, the fibroblast population was subset for further sub-clustering analysis. Sub-clustering was performed using the first 20 PCs at a resolution of 0.3, leading to the identification of 20 distinct fibroblast subpopulations.

2.10 GO and KEGG pathway enrichment analysis

To elucidate the biological functions of key gene set, Gene Ontology (GO) and Genomes (KEGG) pathway enrichment analyses were conducted using the clusterProfiler (v4.16.0). Gene symbols were converted to Entrez IDs. Enrichment analysis was performed for Biological Process (BP), Cellular Component (CC), and Molecular Function (MF) ontologies in GO, as well as for KEGG pathways. Terms with a Benjamini-Hochberg adjusted p-value ($p_{\text{adjust}} < 0.05$) were considered statistically significant. The top enriched terms were visualized using bar plots (for GO) and dot plots (for KEGG).

2.11 Screening of candidate disease genes using LASSO regression

To identify core candidate genes from the gene intersection set, least absolute shrinkage and selection operator (LASSO) regression analysis was performed. The expression matrix of the genes across all samples was randomly divided into a training set (70%) and a validation set (30%) using a random seed. A binomial LASSO regression model was fitted on the training set using the glmnet R package. Ten-fold cross-validation was applied to determine the optimal penalty parameter. The λ_{1se} criterion, which provides the most parsimonious model within one standard error of the minimum cross-validation error, was selected to prevent overfitting.

2.12 Validation via ROC curve analysis

Calculate the area under the curve (AUC) for the test set and validation set to evaluate the accuracy of the model's prediction of positive and negative events. Use an AUC value higher than 0.6 as the threshold for effectively distinguishing between positive and negative samples. The visualization is achieved through the ggplot2 package (version 4.0.1) for plotting and presentation.

3 Results

3.1 Identification of Significantly DEGs in Diabetic Nephropathy

Differential expression analysis was performed on the transcriptome data from DN samples versus normal control samples, yielding 505 significantly differentially expressed genes (Figure 1), including 123 up-regulated genes and 383 down-regulated genes, which were identified as candidate genes potentially associated with DN.

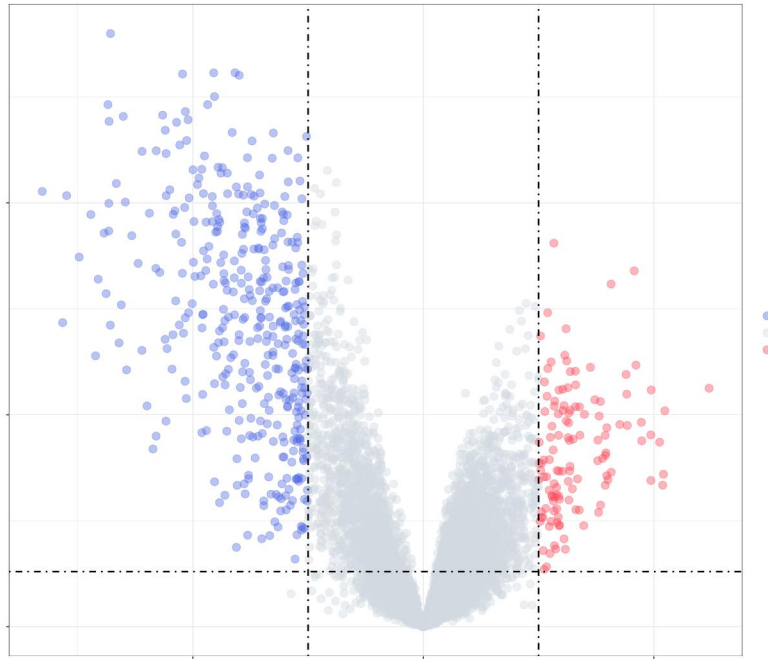


Figure 1 Differentially Expressed Genes between DN Samples and Controls

These DEGs were searched across all 3 categories (biological process, cellular component, and molecular function) in the GO database, and were distinguished between groups using purple, green, and red respectively. The enriched functions of the differentially expressed genes in each category are shown in Figure 2, presenting the top 10 most significant results (p value < 0.05) for each category. Additionally, all KEGG functional enrichment results of the differentially expressed genes are shown in Figure 3. The changes in p value are indicated by the color gradient from light to dark.



Figure 2 The GO Enrichment Results in Each Category

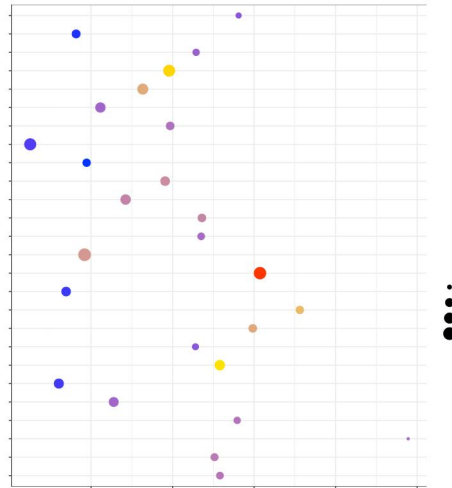


Figure 3 KEGG Enrichment Results for Identified DEGs

3.2 Identification of Co-expression Modules Significantly Associated with Diabetic Nephropathy

3.2.1 WGCNA Network Construction and Module Detection

Given that DEG analysis only captures individual gene changes, we next employed WGCNA to identify gene co-expression modules and explore their systematic association with DN. To investigate the systematic changes in gene expression patterns associated with diabetic nephropathy (DN), WGCNA was performed on the microarray dataset. To construct a scale-free co-expression network, soft-thresholding powers ranging from 1 to 30 were evaluated. A soft-thresholding value of $\beta = 10$ was selected based on the scale-free topology criterion (Figure 4).

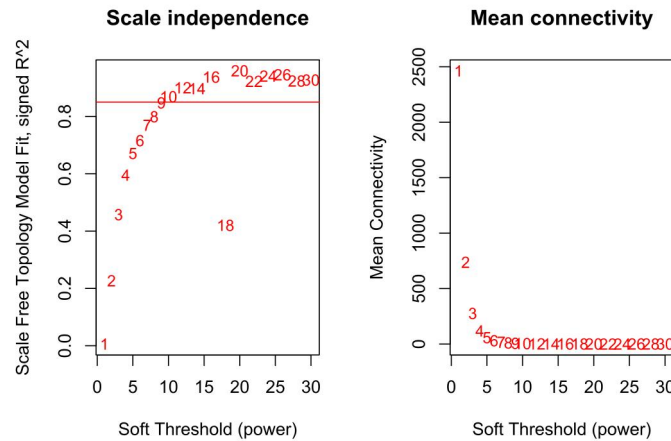


Figure 4 Soft-thresholding power selection for network construction. The left panel shows the scale-free topology model fit index (R^2) as a function of soft-thresholding power (β), and the right panel displays the mean connectivity.

3.2.2 Module-Trait Relationship Analysis

To identify modules significantly associated with DN, the correlation between module eigengenes (MEs) and clinical traits was analyzed, including glomerular and tubular tissues from both diabetic and control samples. The module-trait relationship heatmap revealed distinct correlation patterns (Figure 5).

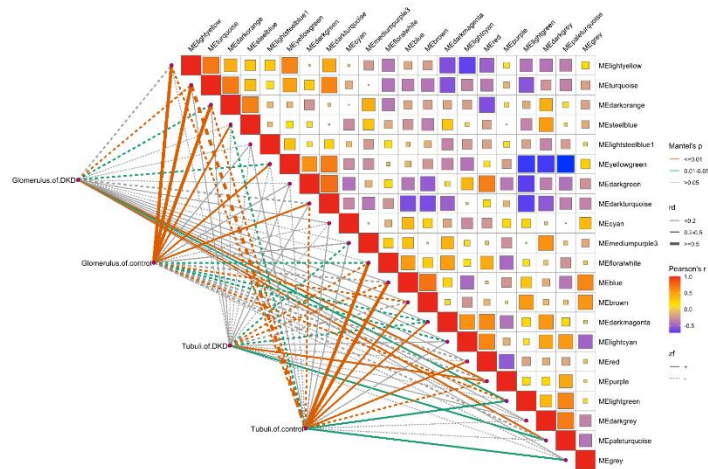


Figure 5 Correlation of Module Eigengenes (MEs) with The Clinical Trait of Diabetic Nephropathy

Notably, the turquoise and darkorange modules showed strong positive correlations with control glomerulus but negative correlations with control tubule (CT), suggesting tissue-specific expression patterns in normal kidney. However, the purple module exhibited positive correlations with diabetic kidney glomerulus (DKG) and tubule (DKT), while showing negative correlations with control tissues, indicating its disease-specific association. Additionally, the blue module displayed opposite correlation trends between diabetic and control tubule samples (Figure 5).

These results suggest that turquoise and darkorange modules are associated with normal tissue specificity, whereas purple and blue modules are linked to DN pathogenesis.

3.2.3 Identification of Key Modules and Hub Genes

Based on the module-trait correlation analysis and scatter plot validation (Figure 6), three key modules were selected for further analysis: turquoise, darkorange, and purple.

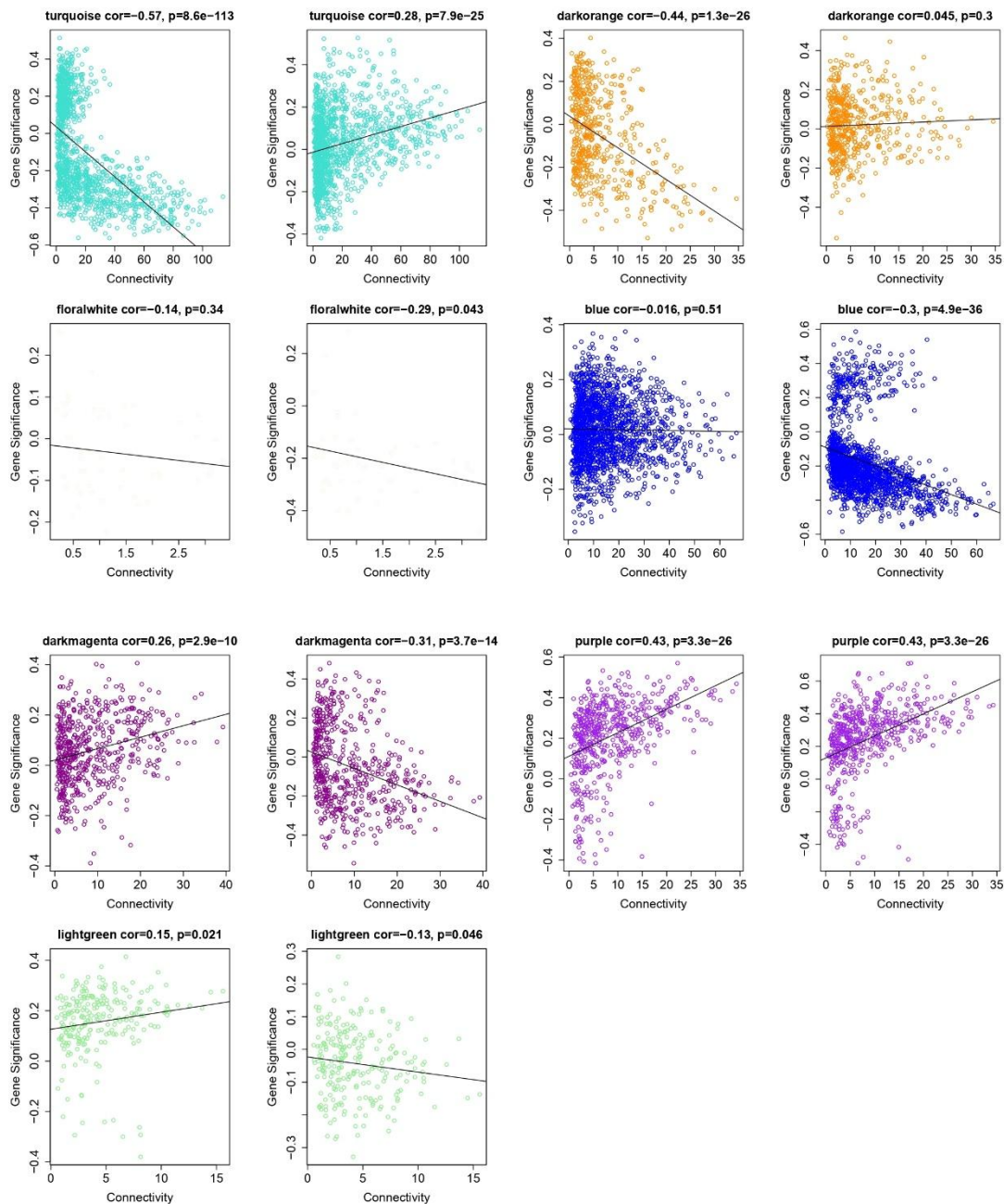


Figure 6 Correlation scatter plots of gene significance (GS) versus module membership (MM)

To identify the most representative genes within these modules, module membership (kME) was calculated for each gene. Genes with $kME > 0.7$ were defined as hub genes. Consequently, 362 hub genes were identified in the turquoise module, 124 in the darkorange module, and 170 in the purple module. These hub genes represent the core components of their respective modules and may play critical roles in DN development or normal kidney function maintenance.

3.3 Identification of Significantly DEGs in Diabetic Retinopathy

After completing the quality control and eliminating all genes that did not show expression in any of the samples, DESeq2 was used to conduct both differential expression analysis and normalization on the RNA-sequencing count data. The results of the differential expression analysis between the diabetic retinopathy (DR) samples and the normal control samples indicated that there were 969 genes with significant differential expression (Figure 7).

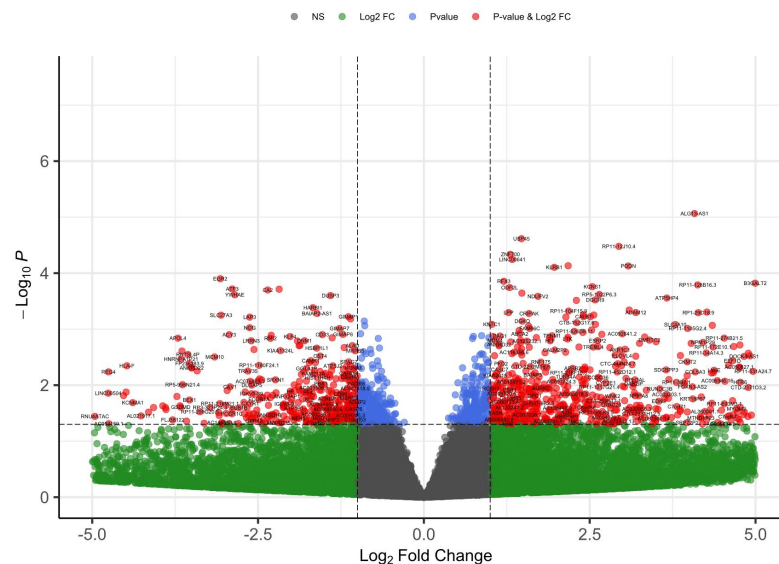


Figure 7 Differentially Expressed Genes between DR Samples and Controls

To gain insights into the biological processes underlying these transcriptional changes, functional enrichment analysis was performed. At the significance threshold of $p < 0.05$, 23 GO terms were significantly enriched. The top 10 GO categories ranked by enrichment significance are presented in Figure 8.

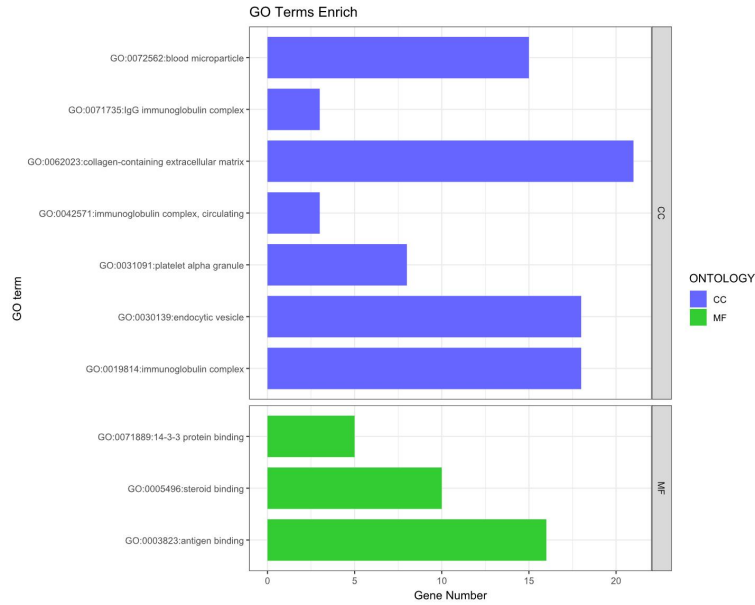


Figure 8 Top 10 GO Enrichment Results in Each Category

The top 10 functional categories were mainly related to immune response, platelet activation, and extracellular matrix remodeling (Figure 8). Among them, immunoglobulin complex (GO:0019814) and antigen binding (GO:0003823) showed the most significant enrichment, suggesting strong humoral immunity activation in the diabetic retina. Terms such as blood microparticle (GO:0072562) and platelet alpha granule (GO:0031091) further highlighted platelet-driven coagulation and vascular dysfunction. The collagen-containing extracellular matrix category (GO:0062023) included 21 genes, including *COL1A1*, *COL1A2* and *FN1*, consistent with the fibrogenic features of proliferative diabetic retinopathy.

This enrichment pattern indicates that the pathological process of DR involves not only changes in the vascular basement membrane, such as collagen components, and a tendency toward microthrombosis, but also is closely related to the activation of local humoral immunity and inflammatory responses.

3.4 Identification of Co-expression Modules Significantly Associated with Diabetic Retinopathy

3.4.1 WGCNA Network Construction and Module Detection

Given that individual DEG analysis may overlook coordinated gene expression patterns, WGCNA was conducted on the 969 DEGs to identify functionally coherent gene modules. Network topology analysis suggested a soft-thresholding power of $\beta = 9$ as the optimal parameter achieving scale-free topology (Figure 9).

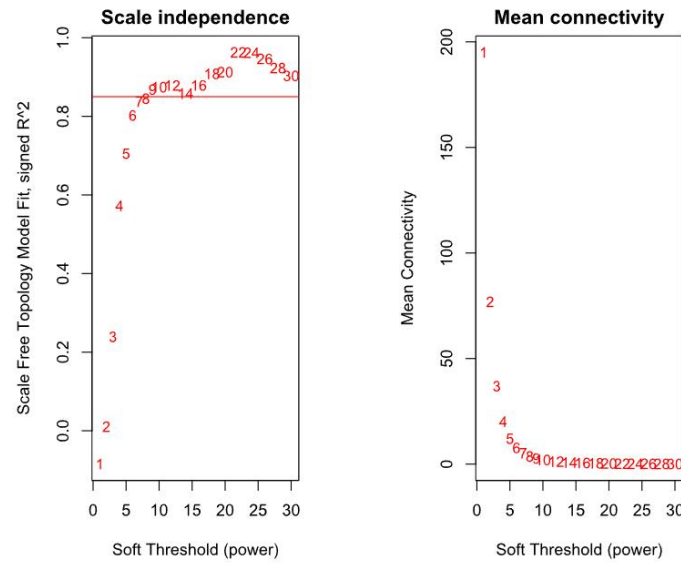


Figure 9 Soft-thresholding power selection for network construction. The left panel shows the scale-free topology model fit index (R^2) as a function of soft-thresholding power (β), and the right panel displays the mean connectivity.

Module-trait correlation analysis revealed distinct specific associations between different gene co-expression modules and clinical phenotypes (Figure 10). Among them, the turquoise module showed an extremely strong positive correlation with diabetic retinopathy (DR) status, while the brown module was highly positively correlated with diabetes mellitus (DM) status. The grey module exhibited a strong negative correlation with the control (CTR) group.

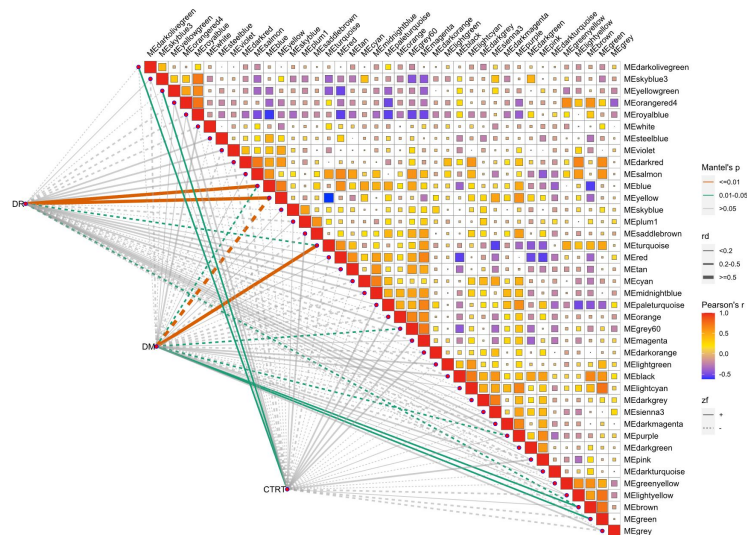


Figure 10 The Module-Trait Correlation Heatmap

These findings suggest that the turquoise module captures the transcriptional programs specific to diabetic retinopathy (DR), whereas the brown module reflects the broader gene expression changes associated with diabetes.

To validate module-phenotype associations, correlations between gene significance (GS) and module membership (MM) were analyzed (Figure 11). Strong GS-MM correlations were observed for both the turquoise and brown modules, supporting their respective biological relevance to diabetic retinopathy (DR) and diabetes. Accordingly, the turquoise and brown modules were selected as core disease-associated modules for subsequent hub gene identification.

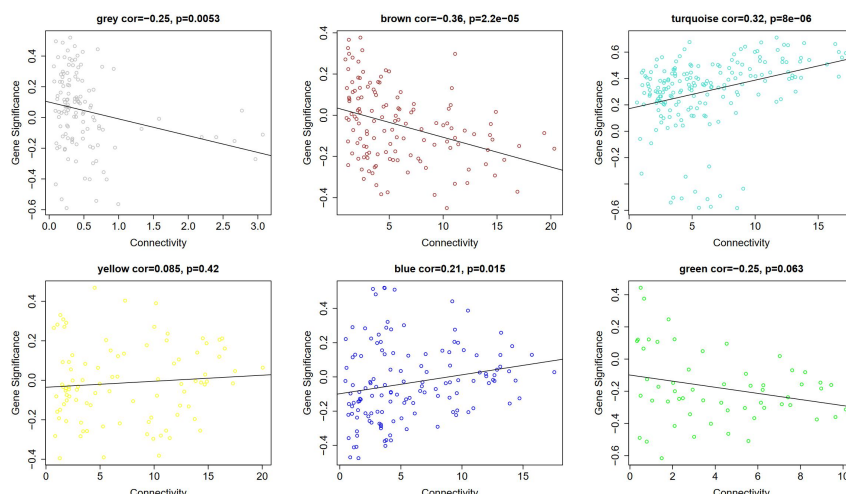


Figure 11 GS-MM correlation scatter plots

To delineate the internal architecture of each core module, KME values were calculated. With a threshold of $KME > 0.7$, 80 hub genes were identified in the turquoise module and 64 in the brown module. These hub genes form the regulatory cores of their respective modules and are likely to serve critical roles in the pathogenesis and progression of DR.

3.5 Analysis of Cellular Composition under Obesity Diet-Induced Oxidative Stress

Single-cell RNA sequencing was performed on liver tissues from HFHF diet-fed aged rats and age-matched control rats. Following Cell Ranger processing, the initial dataset comprised 221,377 genes across 163,844 cells.

Stringent quality control was applied to remove low-quality cells and noise, cells with extreme gene counts or high mitochondrial content were excluded, and genes expressed in fewer than 0.5% of cells or with zero expression were filtered. These criteria yielded a high-quality dataset of 11,320 genes and 150,934 cells for downstream analysis.

Unsupervised clustering was performed following standard workflow. Variable gene selection, PCA dimensionality reduction, and graph-based clustering using $\text{resolution} = 0.7$, partitioned the cells into 34 initial clusters.

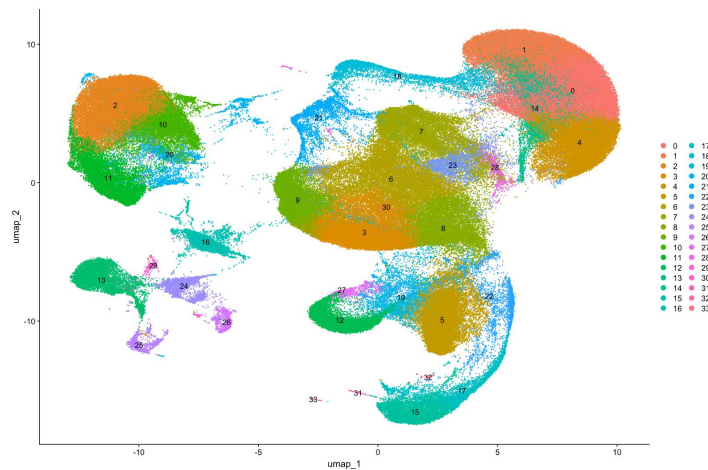


Figure 12 UMAP Visualization

Marker gene analysis identified cluster-specific signatures, which were integrated with reference-based annotation (SingleR) and manual curation using established lineage markers. This approach resolved eight major cell populations: hepatocytes, fibroblasts, endothelial cells, macrophages, T cells, B cells, NK cells, and beta cells.

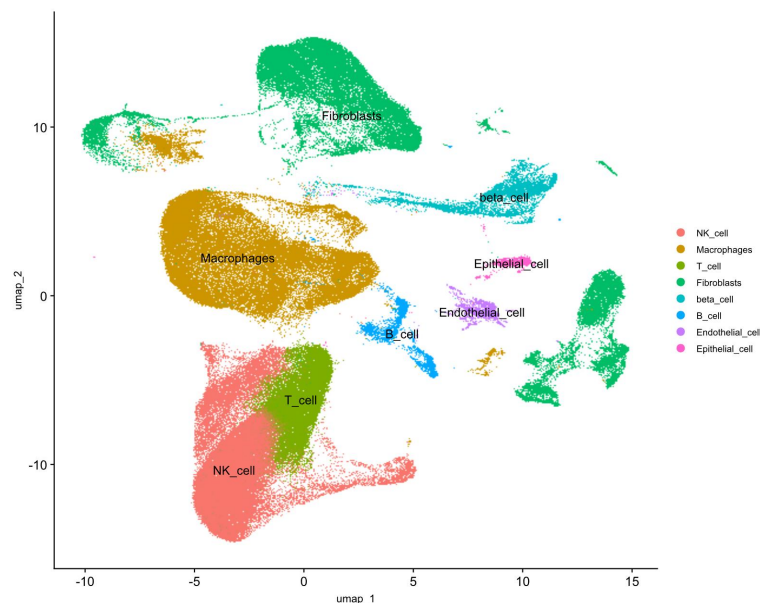


Figure 13 Dot Plot of Top Marker Genes

Comparative analysis revealed marked alterations in cellular composition under HFHF diet-induced metabolic stress. Notably, fibroblasts exhibited substantial expansion, suggesting activation of profibrotic programs in response to chronic nutrient overload. The proportional changes in immune cell subsets like macrophages and T cells further indicated low-grade inflammation, consistent with established features of dysfunctional adipose tissue and metabolic liver disease[6,7].

3.6 The Pivotal Role of Fibrosis in Oxidative Stress

Given the prominent fibroblast expansion observed in HFHF livers, this population was subjected to subclustering analysis to dissect its heterogeneity. Re-clustering at resolution 0.3 identified 20 distinct fibroblast subclusters (Figure 14). Differential expression analysis revealed subcluster-specific gene signatures, with notable diversity in extracellular matrix gene expression profiles.

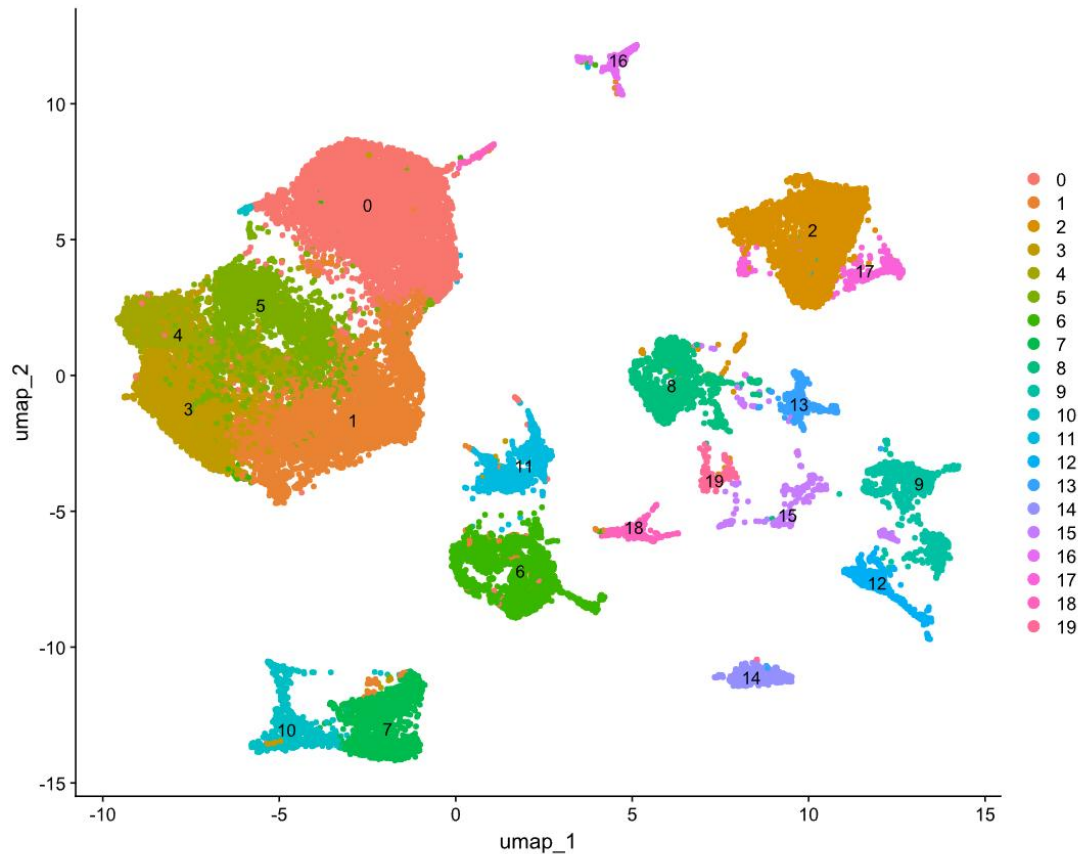


Figure 14 UMAP visualization of subclusters identified at resolution 0.3

This heterogeneity aligns with the recognized complexity of fibrogenesis in metabolic disease, where fibroblasts adopt diverse phenotypes ranging from matrix-producing myofibroblasts to inflammatory fibroblast^[8]. The subcluster distribution patterns suggest that HFHF diet induces not merely fibroblast proliferation but rather qualitative reprogramming toward profibrotic states.

The expansion of specific fibroblast subpopulations under HFHF conditions supports the central role of fibrosis as a convergent pathway linking nutrient excess, oxidative stress, and tissue damage. These findings mirror observations in diabetic complications across multiple organs, where hyperglycemia and lipotoxicity drive extracellular matrix deposition through both direct fibroblast activation and indirect mechanisms involving immune cell recruitment and endothelial dysfunction^[8].

3.7 Linking Endothelial Cells to Diabetic Microvascular Complications and Oxidative Stress for Key Candidate Gene Prediction

To identify conserved mechanisms linking endothelial dysfunction to diabetic microvascular complications, we performed cross-dataset integration. Endothelial cell marker genes derived from the HFHF scRNA-seq dataset were intersected with DEGs from DN and DR transcriptome datasets. This approach yielded 34 genes common to endothelial cells and DN, 2 gene shared with DR, and 36 unique intersection genes overall (Figure 15).

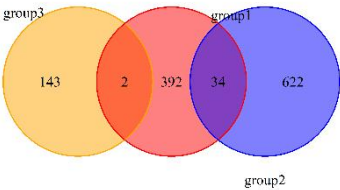


Figure 15 Venn diagram showing overlap between HFHF liver endothelial markers and differentially expressed genes from DN and DR.

The 36-gene intersection set was subjected to functional characterization. Gene Ontology enrichment analysis yielded 228 significant terms, predominantly encompassing immune-mediated mechanisms (lymphocyte-mediated immunity, humoral immune response, complement activation), extracellular matrix remodeling (collagen-containing ECM, glycosaminoglycan binding), and vascular structural components (blood microparticle, basement membrane) (Figure 16). These findings indicate that the conserved endothelial gene signature intersects immune-inflammatory activation and matrix remodeling, suggesting that diabetic microvascular complications arise from coordinated dysregulation of these interconnected processes.



Figure 16 Top enriched GO terms

Moreover, KEGG pathway analysis identified 10 significantly enriched pathways, with *Staphylococcus aureus* infection and phagosome showing the highest enrichment significance, alongside complement and coagulation cascades (Figure 17). These findings highlight immune response activation and pathogen recognition mechanisms as prominent features of the conserved endothelial signature in diabetic complications.

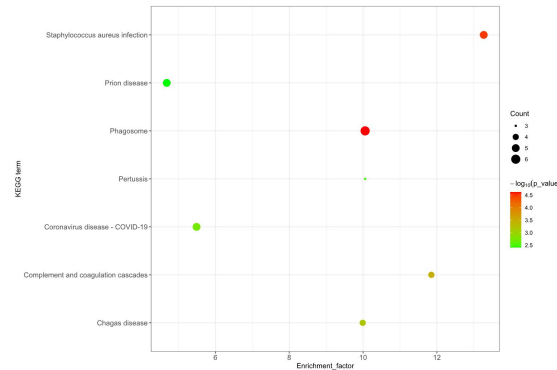


Figure 17 Dot plot of Top 3 KEGG results

3.8 LASSO regression for evaluating important candidate biomarkers

To prioritize candidate genes from the 36-gene intersection set, LASSO regression was applied. Cross-validation determined the optimal lambda value (lambda1.se) as the indicator for screening important features (Figure 18). After re-modeling the data based on the optimal lambda value, four candidate genes were obtained: *NOTCH1*, *CELF2*, *EMP3*, and *VGLL4*.

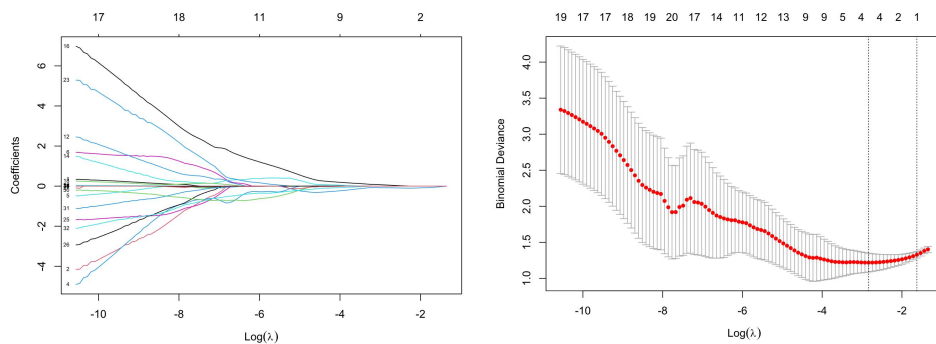


Figure 18 LASSO Regression for Feature Selection of Endothelial Dysfunction
Genes: (Left) LASSO Coefficient Path, (Right) Cross-Validation Curve

3.9 Receiver Operating Characteristic Curve (ROC) Validation

To assess the reliability of these four candidate genes, the data samples were divided into a training set and a test set in a 7:3 ratio. At the same time, the ROC

curves for the four genes were calculated and plotted to examine the false positive probability. As shown in Figure 19, the area under the curve (AUC) of both the training set and the test set of the model is above 0.6. This indicates that the above four genes can be preliminarily regarded as candidate biomarkers for subsequent studies on DN and DR treatment.

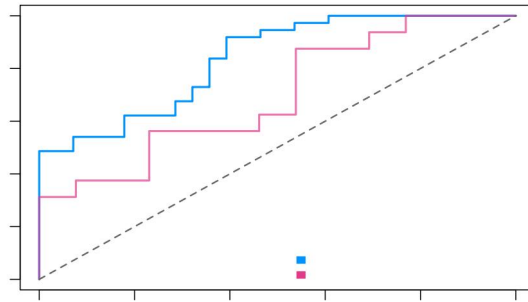


Figure 19 The AUC calculation results of training sets and test sets divided in different proportions

4 Discussion

4.1 Single-cell dynamics and fibroblast expansion

Diabetic nephropathy (DN) and diabetic retinopathy (DR) share highly similar pathophysiological mechanisms, characterized primarily by thickening of the microvascular basement membrane, pericyte loss, and widespread endothelial dysfunction^[9].

In the single-cell analysis of this study, the high-fat, high-fructose (HFHF) diet-induced rat model exhibited a pronounced expansion of fibroblasts, a change that closely mirrors the fibrotic processes triggered by diabetes^[8]. **More importantly, an in-depth examination of endothelial cell (EC) subpopulations revealed that their signature marker genes were significantly differentially expressed in both human DN and DR datasets.** These findings indicate that endothelial cells are not only direct targets of hyperglycemia-induced injury^[11] but also central signal transducers that mediate microvascular damage shared by the kidney and retina^[10]. Consistent with this, the endothelial cells highlighted in our analysis exhibited pronounced transcriptomic alterations under diabetic conditions, and their marker genes showed substantial overlap with the differentially expressed genes identified in human DN and DR. Together, these findings provide strong evidence that endothelial dysfunction constitutes a shared pathological foundation linking these two complications.

4.2 Mechanistic Roles of the Hub Genes in ROS-Mediated Injury

After identifying endothelial cells (ECs) at the single-cell level as the central signal transducers mediating microvascular injury in both the kidney and retina, we further employed an integrative multi-omics strategy to pinpoint the key molecular drivers of this pathological process. Specifically, we intersected EC-specific marker genes identified from the rat single-cell sequencing dataset with the shared differentially expressed genes (DEGs) derived from human DN (GSE30122) and DR (GSE221521) transcriptomic datasets. **Following stringent filtering, *NOTCH1*, *CELF2*, *EMP3*, and *VGLL4* emerged as the most biologically meaningful core candidate genes.**

The analysis identified *NOTCH1* as a top candidate gene, and its role in microvascular complications is well-supported by emerging evidence in both renal and retinal contexts. In the kidney, *NOTCH1* signaling is significantly activated under high-glucose conditions, where it has been shown to augment VEGF expression and suppress nephrin, leading to podocyte loss and the subsequent development of proteinuria^[13]. Conversely, in the diabetic retina, *NOTCH1* activation

(mediated by ligands such as *JAGGED1*) disrupts endothelial adherens junctions by causing the dissociation of VE-cadherin from β -catenin, thereby triggering vascular leakage and macular edema^[12-14].

Among the core genes identified, *CELF2*, a key RNA-binding protein, has attracted increasing attention for its potential role in diabetic complications. Our analysis shows that *CELF2* exhibits marked dysregulation in the affected tissues of both DN and DR, suggesting its involvement in the shared pathogenic mechanisms underlying these conditions^[15]. *CELF2* has been reported to regulate cellular functions primarily through controlling alternative splicing of target genes^[16]. Specifically, *CELF2* acts as a central regulatory factor by modulating the splicing and processing of key pro-inflammatory cytokines and chemokines, including *CXCL2*, *CXCL10*, *IL-1A*, and *IL-6*, thereby contributing to endothelial injury associated with diabetic conditions^[15]. Within the shared pathological context of DN and DR, dysregulation of *CELF2* may disrupt the splicing machinery in endothelial cells, leading to a state of “splicing stress.”^[17] Such perturbations not only exacerbate the release of pro-inflammatory cytokines such as *IL-6* but may also impair the endothelial capacity to eliminate reactive oxygen species^[18-19]. To sum up, *CELF2* serves as a common molecular denominator in DN and DR, bridging abnormal alternative splicing with sustained inflammatory activation and promoting the shift of endothelial cells from quiescence to a pro-inflammatory, pro-angiogenic state central to both diseases.

In the recent study, the core gene *EMP3* was identified and is closely associated with the process of endothelial–mesenchymal transition (EndMT). EndMT is a complex cellular reprogramming event in which endothelial cells progressively lose their characteristic junctional proteins and acquire mesenchymal marker, thereby directly contributing to fibrotic remodeling.

In the diabetic microvascular milieu, high glucose and its metabolic by-products have been shown to activate multiple signaling cascades—including the *TGF- β /SMAD*, Notch, and Wnt/ β -catenin pathways—that collectively drive the EndMT^[20]. Notably, *EMP3*, as a transmembrane glycoprotein, is recognized as an important regulator of membrane stability and cell adhesion. **Within the pathological contexts of DN and DR, the aberrant upregulation of *EMP3* reflects a loss of endothelial homeostasis.** In the kidney, this phenotypic shift contributes to the collapse of glomerular capillary loops and the development of glomerulosclerosis; in the retina, endothelial-derived fibroblast-like cells constitute a major cellular source of the pathological fibrovascular membranes that ultimately lead to tractional retinal detachment^[21,22]. Metabolic dysregulation also contributes substantially to EndMT. Recent studies show that glucose-derived metabolites such as acetate can enhance

TGF- β signaling through enzymes like *ACC2* under hyperglycemic conditions, thereby accelerating EndMT^[23]. Consistent with this mechanism, endothelial cells in the HFHF-induced diabetic model experience oxidative stress-mediated injury and undergo *EMP3*-driven EndMT, shifting toward a pro-fibrotic phenotype. This transition aligns with the marked increase in fibroblast abundance observed in our single-cell sequencing data.

Among the four hub genes identified in this study, *VGLL4* (Vestigial-like family member 4) exerts a unique biological function in maintaining microvascular endothelial homeostasis. As a natural antagonist of the Hippo signaling pathway, *VGLL4* primarily functions by competitively binding to *TEAD* transcription factors via its TDU domains, thereby antagonizing the transcriptional activity of *YAP* (Yes-associated protein)^[24,25]. Extensive evidence has confirmed that the Hippo pathway serves as a "molecular brake" in regulating angiogenesis and endothelial cell proliferation^[26]. Under hyperglycemic metabolic derangement, endothelial YAP is frequently overactivated, driving the expression of pro-angiogenic factors such as VEGF and leading to pathological neovascularization. In diabetic retinopathy (DR), this aberrant YAP activation is a core mechanism of proliferative vitreoretinopathy, characterized by the disordered expansion of fragile new vessel^[27]. When *VGLL4* expression is suppressed, the enhanced stability of the YAP/TEAD complex significantly increases cellular sensitivity to high-glucose-induced oxidative stress and inflammatory cytokines, further accelerating microvascular regression and basement membrane thickening^[28,29].

4.3 Clinical Value and the ROS Nexus

The surge of Reactive Oxygen Species (ROS) is widely recognized as the primary catalyst for diabetic microvascular damage, serving as a critical bridge between hyperglycemic metabolic derangement and pathological tissue remodeling. Our multi-omics integration suggests that the four identified hub genes—*NOTCH1*, *CELF2*, *EMP3*, and *VGLL4*—play decisive roles in modulating both the generation of ROS and the subsequent downstream pathological signaling pathways.

Specifically, a significant positive feedback loop exists between **NOTCH1** signaling and oxidative stress. Hyperglycemia-induced ROS accumulation triggers the activation of NOTCH signaling; in turn, activated NOTCH1 has been shown to impair the function of endothelial nitric oxide synthase (eNOS), leading to abnormal nitric oxide (NO) production. This imbalance further amplifies local oxidative stress and disrupts the integrity of the vascular barrier^[30,31].

Simultaneously, **CELF2**-mediated alternative splicing dysregulation and **EMP3**-driven endothelial-to-mesenchymal transition (EndMT) act as key ROS-induced stress effectors. These processes exacerbate local pro-inflammatory cytokine release and facilitate the progression of tissue fibrosis, respectively^[32,33]. Furthermore, the downregulation of **VGLL4**, which serves as a "homeostatic brake," signifies the loss of an intrinsic protective shield against ROS-induced injury. This loss of protection renders the microvascular system more susceptible to regression and apoptosis under sustained metabolic pressure^[27].

From a clinical translation standpoint, the diagnostic model built around these four hub genes showed strong and consistent performance in several independent human transcriptomic cohorts, with AUC values remaining above 0.8 across datasets. The reproducibility of these results in different tissues and study populations suggests that these genes may represent common molecular mechanisms underlying "renal-retinal syndrome."

While traditional clinical assessments often treat diabetic nephropathy (DN) and diabetic retinopathy (DR) as isolated entities, my findings suggest that this specific panel of molecular markers enables a "single-marker, dual-diagnosis" approach. This allows for the simultaneous monitoring of microvascular damage in both the kidney and the retina through a unified molecular assessment. Such an approach not only provides a robust tool for early screening and precision monitoring of disease progression but also establishes a theoretical foundation for developing "dual-benefit" therapies that target the critical nodes downstream of oxidative stress.

5 Conclusion

In summary, this study provides an integrative framework illustrating how diet-induced oxidative stress contributes to the development of DN and DR. By combining the single-cell transcriptome data of the high-fat and high-sugar-induced oxidative stress model with the overall human transcriptome data, it was found that endothelial dysfunction is the central and shared pathological axis connecting microvascular damage in the kidneys and the retina. The expansion of fibroblasts and the transcriptional reprogramming of endothelial cells under metabolic stress further support the role of fibrosis and vascular instability as key downstream consequences of ROS accumulation.

Through cross-dataset integration and LASSO regression, four hub genes—*NOTCH1*, *CELF2*, *EMP3*, and *VGLL4*—were identified as core candidates mediating oxidative stress-related endothelial injury. These genes collectively participate in signaling pathways associated with inflammation, endothelial-to-mesenchymal transition, alternative splicing regulation, and Hippo pathway modulation, forming a coordinated regulatory network that bridges hyperglycemia-induced oxidative imbalance with microvascular structural remodeling.

Collectively, these results support the view that DN and DR represent two organ-specific manifestations of a common oxidative stress-driven endothelial injury program. Future studies can focus on experimental validation of the functional roles of these hub genes in vivo and in vitro, as well as prospective clinical investigations to determine whether targeting these shared molecular pathways can achieve simultaneous protection of renal and retinal function. Such efforts may pave the way for the development of dual-organ therapeutic strategies in diabetic patients.

6 Reference

- [1].Woroniecka, K., Park, A., Mohtat, D., Thomas, D., Pullman, J., & Susztak, K. (2011). Transcriptome Analysis of Human Diabetic Kidney Disease. *Diabetes*, 60, 2354 - 2369. <https://doi.org/10.2337/db10-1181>.
- [2].Na J, Sweetwyne MT, Park AS, Susztak K et al. Diet-Induced Podocyte Dysfunction in Drosophila and Mammals. *Cell Rep* 2015 Jul 28;12(4):636-47. PMID: 26190114 <https://doi.org/10.1016/j.celrep.2015.06.056>
- [3].Lam JD, Oh DJ, Wong LL, Amarnani D et al. Identification of RUNX1 as a Mediator of Aberrant Retinal Angiogenesis. *Diabetes* 2017 Jul;66(7):1950-1956. PMID: 28400392 <https://doi.org/10.2337/db16-1035>
- [4].Zhao Y, Wang QY, Zeng LT, Wang JJ et al. Long-Term High-Fat High-Fructose Diet Induces Type 2 Diabetes in Rats through Oxidative Stress. *Nutrients* 2022 May 24;14(11). PMID: 35683981 <https://doi.org/10.3390/nu14112181>
- [5].Ma S, Sun S, Geng L, Song M et al. Caloric Restriction Reprograms the Single-Cell Transcriptional Landscape of Rattus Norvegicus Aging. *Cell* 2020 Mar 5;180(5):984-1001.e22. PMID: 32109414 <https://doi.org/10.1016/j.cell.2020.02.008>
- [6].De Fano, M., Bartolini, D., Tortoioli, C., Vermigli, C., Malara, M., Galli, F., & Murolo, G. (2022). Adipose Tissue Plasticity in Response to Pathophysiological Cues: A Connecting Link between Obesity and Its Associated Comorbidities. *International Journal of Molecular Sciences*, 23(10), 5511. <https://doi.org/10.3390/ijms23105511>
- [7].Hou, Y., Ding, W., Wu, P. et al. Adipose-derived stem cells alleviate liver injury induced by type 1 diabetes mellitus by inhibiting mitochondrial stress and attenuating inflammation. *Stem Cell Res Ther* 13, 132 (2022). <https://doi.org/10.1186/s13287-022-02760-z>
- [8].Tuleta, I., & Frangogiannis, N. G. (2021). Diabetic fibrosis. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1867(4), 166044. <https://doi.org/10.1016/j.bbadis.2020.166044>
- [9].Kanwar, Y. S., Sun, L., Xie, P., Liu, F. Y., & Chen, S. (2011). A glimpse of various pathogenetic mechanisms of diabetic nephropathy. *Annual review of pathology*, 6, 395–423. <https://doi.org/10.1146/annurev.pathol.4.110807.092150>
- [10].Li, Z., Wang, Y. Biomarker signatures distinguish diabetic retinopathy and diabetic nephropathy. *Diabetol Int* 16, 735–748 (2025). <https://doi.org/10.1007/s13340-025-00842-w>

- [11].Wang E, Feng B, Chen S, Su Z, Chakrabarti S (2024) Differential microvascular endothelial cell responses in the retina in diabetes compared to the heart and kidneys, a spatial transcriptomic analysis. PLOS ONE 19(12): e0310949. <https://doi.org/10.1371/journal.pone.0310949>
- [12].Chun-Liang Lin, Feng-Sheng Wang, Yen-Chen Hsu, Cheng-Nan Chen, Min-Jen Tseng, Moin A. Saleem, Pey-Jium Chang, Jeng-Yi Wang; Modulation of Notch-1 Signaling Alleviates Vascular Endothelial Growth Factor–Mediated Diabetic Nephropathy. Diabetes 1 August 2010; 59 (8): 1915–1925. <https://doi.org/10.2337/db09-0663>
- [13].Lin, C. L., Wang, F. S., Hsu, Y. C., Chen, C. N., Tseng, M. J., Saleem, M. A., Chang, P. J., & Wang, J. Y. (2010). Modulation of notch-1 signaling alleviates vascular endothelial growth factor-mediated diabetic nephropathy. Diabetes, 59(8), 1915–1925. <https://doi.org/10.2337/db09-0663>
- [14].Miloudi, K., Oubaha, M., Ménard, C., Dejda, A., Guber, V., Cagnone, G., Wilson, A., ... Sapienza, P. (2019). NOTCH1 signaling induces pathological vascular permeability in diabetic retinopathy. Proceedings of the National Academy of Sciences of the United States of America, 116(10), 4538–4547. <https://www.jstor.org/stable/26683126>
- [15].Nuerdebek, D., Yao, L., Guo, L., Li, J., Jia, H., & Nan, Y. (2025). Gene expression profiling of human umbilical vein endothelial cells overexpressing CELF2 as diagnostic targets in diabetes-induced erectile dysfunction. Frontiers in molecular biosciences, 12, 1596534. <https://doi.org/10.3389/fmolb.2025.1596534>
- [16].Ajith, S., Gazzara, M. R., Cole, B. S., Shankarling, G., Martinez, N. M., Mallory, M. J., & Lynch, K. W. (2016). Position-dependent activity of CELF2 in the regulation of splicing and implications for signal-responsive regulation in T cells. RNA biology, 13(6), 569–581. <https://doi.org/10.1080/15476286.2016.1176663>
- [17].Nutter, C. A., Jaworski, E., Verma, S. K., Perez-Carrasco, Y., & Kuyumcu-Martinez, M. N. (2017). Developmentally regulated alternative splicing is perturbed in type 1 diabetic skeletal muscle. Muscle & nerve, 56(4), 744–749. <https://doi.org/10.1002/mus.25599>
- [18].González, P., Lozano, P., Ros, G., & Solano, F. (2023). Hyperglycemia and Oxidative Stress: An Integral, Updated and Critical Overview of Their Metabolic Interconnections. International Journal of Molecular Sciences, 24(11), 9352. <https://doi.org/10.3390/ijms24119352>
- [19].González, P., Lozano, P., & Ros, G. (2023). Hyperglycemia and oxidative stress: An integral, updated and critical overview of their metabolic

- interconnections. *International Journal of Molecular Sciences*, 24(11), 9352. <https://doi.org/10.3390/ijms24119352>
- [20].Kim, R., & Chang, W. (2025). Endothelial-to-Mesenchymal Transition in Health and Disease: Molecular Insights and Therapeutic Implications. *International Journal of Molecular Sciences*, 26(23), 11724. <https://doi.org/10.3390/ijms262311724>
- [21].Jacobs, M. E., de Vries, D. K., Engelse, M. A., Dumas, S. J., & Rabelink, T. J. (2024). Endothelial to mesenchymal transition in kidney fibrosis. *Nephrology Dialysis Transplantation*, 39(5), 752–760. <https://doi.org/10.1093/ndt/gfad238>
- [22].Zeisberg, E. M., Tarnavski, O., Zeisberg, M., Dorfman, A. L., McMullen, J. R., Gustafsson, E., Chandraker, A., Yuan, X., Pu, W. T., Roberts, A. B., Neilson, E. G., Sayegh, M. H., Izumo, S., & Kalluri, R. (2007). Endothelial-to-mesenchymal transition contributes to cardiac fibrosis. *Nature medicine*, 13(8), 952–961. <https://doi.org/10.1038/nm1613>
- [23].Recatalá, C. A., & Santoro, M. M. (2023). The glucose-to-acetate metabolic flux that drives endothelial-to-mesenchymal transition via TGF- β signaling. *Cell metabolism*, 35(7), 1093–1095. <https://doi.org/10.1016/j.cmet.2023.06.006>
- [24].Mickle, M., Adhikary, G., Shrestha, S., Xu, W., & Eckert, R. L. (2021). VGLL4 inhibits YAP1/TEAD signaling to suppress the epidermal squamous cell carcinoma cancer phenotype. *Molecular carcinogenesis*, 60(7), 497–507. <https://doi.org/10.1002/mc.23307>
- [25].Jiao, S., Wang, H., Shi, Z., Dong, A., Zhang, W., Song, X., He, F., Wang, Y., Zhang, Z., Wang, W., Wang, X., Guo, T., Li, P., Zhao, Y., Ji, H., Zhang, L., & Zhou, Z. (2014). A peptide mimicking VGLL4 function acts as a YAP antagonist therapy against gastric cancer. *Cancer cell*, 25(2), 166–180. <https://doi.org/10.1016/j.ccr.2014.01.010>
- [26].Kimura, T. E., Duggirala, A., Smith, M. C., White, S., Sala-Newby, G. B., Newby, A. C., & Bond, M. (2016). The Hippo pathway mediates inhibition of vascular smooth muscle cell proliferation by cAMP. *Journal of molecular and cellular cardiology*, 90, 1–10. <https://doi.org/10.1016/j.yjmcc.2015.11.024>
- [27].Hooglugt, A., van der Stoep, M. M., Boon, R. A., & Huveneers, S. (2021). Endothelial YAP/TAZ Signaling in Angiogenesis and Tumor Vasculature. *Frontiers in oncology*, 10, 612802. <https://doi.org/10.3389/fonc.2020.612802>
- [28].Wei, L., Gao, J., Wang, L., Tao, Q., & Tu, C. (2023). Hippo/YAP signaling pathway: a new therapeutic target for diabetes mellitus and vascular complications. *Therapeutic advances in endocrinology and metabolism*, 14, 20420188231220134. <https://doi.org/10.1177/20420188231220134>

- [29].Hao, G. M., Lv, T. T., Wu, Y., Wang, H. L., Xing, W., Wang, Y., Li, C., Zhang, Z. J., Wang, Z. L., Wang, W., & Han, J. (2017). The Hippo signaling pathway: a potential therapeutic target is reversed by a Chinese patent drug in rats with diabetic retinopathy. *BMC complementary and alternative medicine*, 17(1), 187. <https://doi.org/10.1186/s12906-017-1678-3>
- [30].K. Miloudi, M. Oubaha, C. Ménard, A. Dejda, V. Guber, G. Cagnone, A.M. Wilson, N. Tétreault, G. Mawambo, F. Binet, R. Chidiac, C. Delisle, M. Buscarlet, A. Cerani, S. Crespo-Garcia, K. Bentley, F. Rezende, J. Joyal, F.A. Mallette, J. Gratton, B. Larrivée, & P. Sapiha(2019). NOTCH1 signaling induces pathological vascular permeability in diabetic retinopathy, *Proc. Natl. Acad. Sci. U.S.A.* 116 (10) 4538-4547, <https://doi.org/10.1073/pnas.1814711116>.
- [31].Murea, M., Park, J. K., Sharma, S., Kato, H., Gruenwald, A., Niranjana, T., Si, H., Thomas, D. B., Pullman, J. M., Melamed, M. L., & Susztak, K. (2010). Expression of Notch pathway proteins correlates with albuminuria, glomerulosclerosis, and renal function. *Kidney international*, 78(5), 514 – 522. <https://doi.org/10.1038/ki.2010.172>
- [32].Nuerdebieke, D., Yao, L., Guo, L., Li, J., Jia, H., & Nan, Y. (2025). Gene expression profiling of human umbilical vein endothelial cells overexpressing CELF2 as diagnostic targets in diabetes-induced erectile dysfunction. *Frontiers in molecular biosciences*, 12, 1596534. <https://doi.org/10.3389/fmolb.2025.1596534>
- [33].He, J., Xu, Y., Koya, D., & Kanasaki, K. (2013). Role of the endothelial-to-mesenchymal transition in renal fibrosis of chronic kidney disease. *Clinical and experimental nephrology*, 17(4), 488-497.