Comparison of Kidney Transcriptomic Profiles of Early and Advanced Diabetic Nephropathy Reveals Potential New Mechanisms for Disease Progression

Ying Fan,[1] Zhengzi Yi,[2] Vivette D. D'Agati,[3] Zeguo Sun,[2] Fang Zhong,[2] Weijia Zhang,[2] Jiejun Wen,[1] Ting Zhou,[1] Ze Li,[1] Li He,[1] Qunzi Zhang,[1] Kyung Lee,[2] John Cijiang He,[2,4] and Niansong Wang[1]

To identify the factors mediating the progression of diabetic nephropathy (DN), we performed RNA sequencing of kidney biopsy samples from patients with early DN, advanced DN, and normal kidney tissue from nephrectomy samples. A set of genes that were upregulated at early but downregulated in late DN were shown to be largely renoprotective, which included genes in the retinoic acid pathway and glucagon-like peptide 1 receptor. Another group of genes that were downregulated at early but highly upregulated in advanced DN consisted mostly of genes associated with kidney disease pathogenesis, such as those related to immune response and fibrosis. Correlation with estimated glomerular filtration rate (eGFR) identified genes in the pathways of iron transport and cell differentiation to be positively associated with eGFR, while those in the immune response and fibrosis pathways were negatively associated. Correlation with various histopathological features also identified the association with the distinct gene ontological pathways. Deconvolution analysis of the RNA sequencing data set indicated a significant increase in monocytes, fibroblasts, and myofibroblasts in advanced DN kidneys. Our study thus provides potential molecular mechanisms for DN progression and association of differential gene expression with the functional and structural changes observed in patients with early and advanced DN.

Diabetic nephropathy (DN) is the most common cause of end-stage renal disease in the U.S., and its incidence is rising worldwide despite glycemic and blood pressure control regimens (1). Therefore, elucidating mechanisms that mediate the early stage of DN may help us to identify new targets for better preventive and therapeutic measures.

Genome-wide gene expression profiling can be useful in providing a global picture of the disease pathogenesis and to identify potential new biomarkers and drug targets for DN. Several previous studies examined the transcriptomes of human diabetic kidney samples. The European Renal cDNA Bank (ERCB) consortium examined the gene expression of the tubulointerstitial compartments of European Caucasian patients with DN compared with those from pretransplant donors and minimal change disease by microarray analysis, which identified the expression of genes related to the nuclear factor-κB–driven inflammatory pathway to be highly associated with the progression of DN (2). The ERCB consortium also found that vascular endothelial growth factor A (*VEGFA*) expression was downregulated in the tubulointerstitial compartment of DN, which is associated with interstitial vascular rarefaction (3). Berthier et al. (4) analyzed the transcriptome of glomerular and tubulointerstitial compartments in patients with early DN from a Pima Indian cohort, patients with progressive DN from ERCB, and control subjects without diabetes and identified the Janus kinase-STAT pathway to be upregulated in both glomerular and tubulointerstitial compartments in DN. Woroniecka et al. (5) analyzed the glomerular and tubular microarray data from kidney biopsy samples of patients with DN compared

[1]Department of Nephrology, Shanghai Jiao Tong University Affiliated Sixth People's Hospital, Shanghai, China

[2]Division of Nephrology, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY

[3]Department of Pathology, Columbia University Medical Center, New York, NY

[4]Kidney Center at James J. Peters VA Medical Center, Bronx, NY

Corresponding author: John Cijiang He, cijiang.he@mssm.edu, or Niansong Wang, wangniansong2012@163.com

COMPLICATIONS

with samples from living allograft donors and surgical nephrectomies, which showed the upregulation of RhoA, Cdc42, integrin, and VEGF signaling in the glomerular compartment and inflammation-related pathways in the tubulointerstitial compartment in DN. Nair et al. (6) examined the microarray analysis of the tubulointerstitial compartments samples from the Pima Indian cohort, which included patients with diabetes with normoalbuminuria, microalbuminuria, and macroalbuminuria, and found that the cortical interstitial fractional volume, an index of tubulointerstitial damage, correlated significantly with the transcripts enriched for pathways associated with mitochondrial dysfunction, inflammation, migratory mechanisms, and tubular metabolic functions. In another recent study by Pan et al. (7), microarray analysis of kidney samples from patients with advanced DN and surgical nephrectomies were performed for glomerular transcriptome profiling, which identified SLIT-ROBO GTPase-activating protein 2a (SRGAP2a) as a key gene associated with proteinuria and estimated glomerular filtration rate (eGFR) in patients with DN. Thus, the transcriptomic analyses are of microarray analyses in samples from patients with advanced DN, with the exception of the Pima Indian study. In most of these previous studies, the transcriptomes were profiled in glomerular and tubulointerstitial compartments separately.

Here, we report an RNA sequencing (RNA-seq) analysis of the whole-kidney biopsy samples of patients with early and advanced DN compared with nephrectomy sample tissues from patients without diabetes. We compared gene expression profiles among normal samples, early DN, and advanced DN. We also performed a correlation analysis of genes with renal function (eGFR) and histological parameters in patients with DN. We took advantage of the recently published single-cell RNA-seq (scRNA-seq) data to perform a computational deconvolution analysis to identify the different cell types present in normal and diseased kidneys. Finally, we validated some of the key findings by immunostaining of the kidney tissues from these patients.

RESEARCH DESIGN AND METHODS

Human Kidney Biopsy Sample Collection

A total of 28 patients with biopsy-proven DN hospitalized from January 2015 to December 2016 in Shanghai Jiao Tong University Affiliated Sixth People's Hospital were enrolled in the study. Kidney tissues were collected through ultrasound-guided kidney biopsy after informed consent was obtained, according to the guidelines of the local ethics committee. Samples were quickly frozen in liquid nitrogen and stored at $-80°C$ before use. Patients were divided into two groups—early DN ($n = 6$) and advanced DN ($n = 22$)—on the basis of urinary albumin-to-creatinine ratio (UACR) and renal function (calculated using the MDRD equation) by at least two randomized measurements. Early DN was defined as UACR between 30 and 300 mg/g, eGFR >90 mL/min/1.73 m^2, whereas advanced DN was defined as UACR >300 mg/g,

eGFR <90 mL/min/1.73 m^2. Nine control human kidney samples were obtained from the unaffected portion of tumor nephrectomies. Demographic, blood biochemical characteristic, urine albumin excretion, and kidney function data were collected. The study was approved by the institutional review board at Shanghai Jiao Tong University Affiliated Sixth People's Hospital.

RNA-Seq and Bioinformatics Analysis

RNA-seq was performed on 28 DN and 9 control samples. Total RNA was extracted using TRIzol (Thermo Fisher Scientific). The RNA quality was evaluated by an Agilent 2100 Bioanalyzer. The total RNA sample was treated using a Ribo-Zero Magnetic Gold Kit to deplete rRNA, and cDNA libraries were prepared and sequenced at BGI using a HiSeq 4000 system. The reads with good quality were first aligned to human reference databases including hg19 human genome, exon, and splicing junction segment and a contamination database including ribosome and mitochondria sequences using the STAR alignment algorithm (8). After filtering reads mapped to the contamination database, the reads that were uniquely aligned to the exon and splicing junction segments with a maximal two mismatches for each transcript were then counted as the expression level for the corresponding transcript. After filtering out the transcripts with low reads (<100) across all samples, the read counts were log$_2$ transformed and quantile normalized at an equal global median value to compare the transcription level across samples. Gene expression data were first adjusted by demographic confounders such as age and sex by extracting residuals from the linear regression model. Principal component analysis was then performed to assess the overall sample distribution using adjusted expression data of all the genes. The differential analysis by limma test (9) was carried out to identify significantly dysregulated genes at $P < 0.05$ (for advanced DN vs. early DN, we used a false discovery rate [FDR] <0.05 because of a large amount of differentially expressed genes [DEGs]) and 1.5-fold change, which was then subjected to Gene Ontology (GO) function (10) and pathway (Kyoto Encyclopedia of Genes and Genomes, Ingenuity Pathway Analysis, BioCarta, Naba, Panther, Pathway Interaction Database, Reactome, WikiPathways) enrichment analysis by Fisher exact test. Correlation of gene expression and eGFR and histology scores were done on all DN samples. We determined eGFR-related genes at an FDR <0.05 by Pearson correlation test. Correlation with histology scores was determined by multivariate analysis using linear regression model at $P < 0.05$ for each score. As for deconvolution analysis, CIBERSORT (11) was used to estimate the percentage of each cell type on the basis of the expression value of markers in bulk sequencing samples. A human kidney allograft biopsy single-cell data set was obtained from the Gene Expression Omnibus (GEO) database (GSE109564) and used to identify 16 cell types by Seurat with parameters described in the original article (12). For each cell type, gene expression

values were compared against other cell types. Genes with t test $P < 0.01$ and $\log_2$(fold change) >0.5 were selected as markers for corresponding cell types. For each marker, we calculated the mean expression value in each cell type to construct the base matrix for deconvolution analysis. The same strategy for deconvolution analysis was also performed on the basis of mouse single-cell sequencing data obtained from the GEO database (GSE107585) (13).

Histological Evaluations
Histological analysis of all patients with DN was performed by investigators blinded to the experimental groups. A renal pathologist scored detailed histological features from whole-slide images of kidney biopsy samples stained with hematoxylin-eosin, periodic acid Schiff, and Masson trichrome as described previously (14).

Immunostaining of Kidney Sections
Human kidney biopsy samples from these patients were collected as described under protocols approved by the institutional review board. Biopsy samples included 5 cases of early DN, 44 of advanced DN, and 7 of normal tissue adjacent to tumor nephrectomy samples. Immunostaining was performed in all five samples of early DN, five randomly selected samples of advanced DN, and five normal samples from nephrectomies using specific primary antibodies and biotinylated secondary antibodies (Vector Laboratories Inc.). Staining was revealed with avidin-peroxidase (VECTASTAIN Elite; Vector Laboratories Inc.). Slides were mounted in Aqua Poly/Mount (Polysciences, Inc.) and photographed under an Olympus BX60 microscope with a digital camera. The following antibodies were used: retinol-binding protein 4 (RBP4) (ab133530; Abcam) and GRP1 (ab166987; Abcam). The extent of kidney staining in human biopsy samples was semiquantitatively scored on the basis of the percentage of positive staining area divided by total kidney cortex area in the kidney section for each patient. Immunostaining for immune cells was performed with anti-Mrc1 for macrophages (ab64693; Abcam), CD3 for T cells (A0452; Dako), and CD20 (M0755; Dako) for B cells.

Data and Resource Availability
The RNA-seq data set generated and analyzed in the current study is available from the GEO repository (GSE142025). All other data sets generated during the current study are available from the corresponding authors upon reasonable request. No applicable resources were generated during the current study.

RESULTS

RNA-Seq of Kidney Biopsy Samples From Patients With DN
Kidney biopsy samples from 28 patients with DN (22 advanced and 6 early) and 9 normal nephrectomy samples adjacent to tumors were used for RNA-seq analysis. Baseline characteristics are presented in Supplementary Table 1.

Patients with early DN had an eGFR >90 mL/min/1.73 m^2 and microalbuminuria (UACR <300 mg/g). Patients with advanced DN had either an eGFR <90 mL/min/1.73 m^2 or UACR >300 mg/g. Patients with advanced DN had significantly lower eGFR and serum albumin but higher UACR, serum creatinine, and blood urea nitrogen levels than patients with early DN. Patients with early DN had a similar duration of diabetes and other clinical parameters to patients with advanced DN, suggesting that they were likely to be nonprogressors. In addition, their UACR and renal function in the 2–3-year follow-up from the time of collection of kidney biopsy samples remained stable (data not shown). To minimize the tissue processing (i.e., without any dissections or digestions) that may change the gene expression profiles (15), RNA-seq was performed using the whole-kidney biopsy samples, which contain primary kidney cortices. The principal component analysis of the RNA-seq data showed three distinct clusters (Supplementary Fig. 1), indicating large changes in the overall gene signatures among the three groups.

Comparison of RNA-Seq Data of Kidney Biopsy Samples Between Patients With Early DN and Control Patients
We first analyzed the DEGs in the kidney cells of patients with early DN compared with control patients without diabetes. Figure 1A shows a heatmap of the top 25 upregulated and top 25 downregulated genes in early DN, and Fig. 1B shows a volcano plot of the top 50 DEGs (limma $P < 0.05$ and fold change >1.5 or less than -1.5) (a complete list of DEGs is included in Supplementary Excel File 1). Of note, among the top 25 downregulated genes were transcription factors that are implicated in the pathogenesis of DN (16), such as *EGR1*, *EGR2*, *EGR3*, *JUNB*, *FOS*, *FOSB*, and *ATF3*. *IL6* and *CXCL2*, inflammatory markers associated with kidney injury in DN (17,18), were also included in the top 25 downregulated genes. GO analysis of DEGs between early DN and control samples is shown in Fig. 2. The top GO terms of the upregulated genes in early DN were related to cellular contraction, containing several myosin and actin genes, and to hormonal regulation and visual perception, containing several genes in the retinoic acid pathway (e.g., *RDH8*, *RDH12*, and *RBP4*). Retinoic acid pathway is shown to be renoprotective in several animal models of kidney disease (19,20). Interestingly, glucagon-like peptide 1 receptor (*GLP1R*) expression was also highly upregulated in early DN. The pathway enrichment analysis of combined upregulated and downregulated DEGs is shown in Supplementary Fig. 2. Many of the downregulated genes were related to immune response, suggesting that the inflammatory pathway is suppressed in early and nonprogressive DN compared with control nephrectomy samples. This could be due to low inflammation status in the kidney of the patients with early DN but may be a result of mildly increased inflammation in the nephrectomy samples from the normal tissues surrounding tumor.
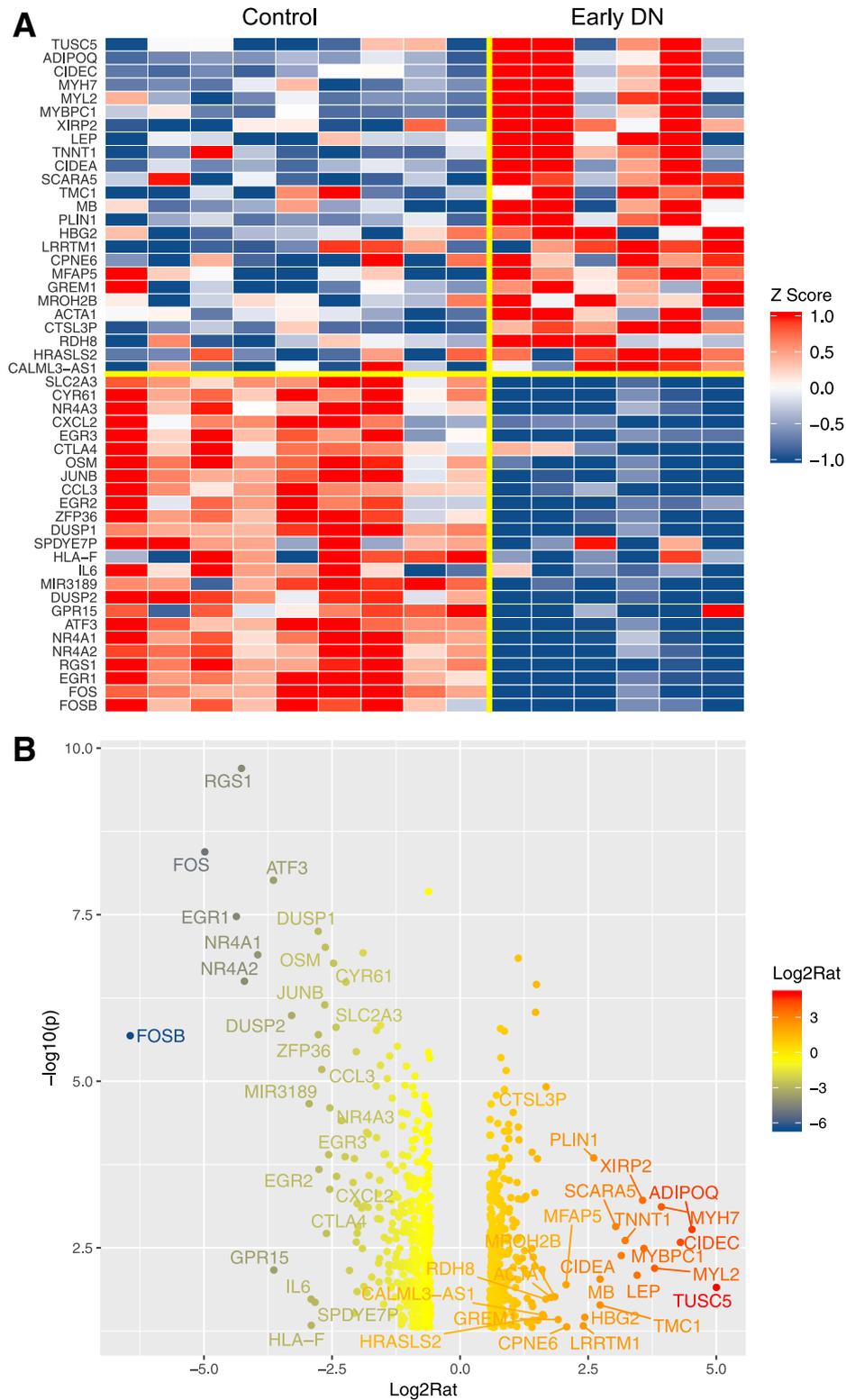
Figure 1—Differential gene expression analysis for early DN vs. nondiabetic control. *A*: Heatmap of the top 50 dysregulated genes (25 upregulated genes and 25 downregulated genes) in early DN samples. Data are *z* normalized for heatmap visualization. Each column represents an individual sample from the control or early DN group. *B*: Volcano plot of dysregulated genes at limma $P < 0.05$ and fold change >1.5 or less than -1.5. The top 50 DEGs are labeled on the plot. Log2Rat, $\log_2$ ratio; $-\log 10(p)$, $-\log_{10} P$ value.
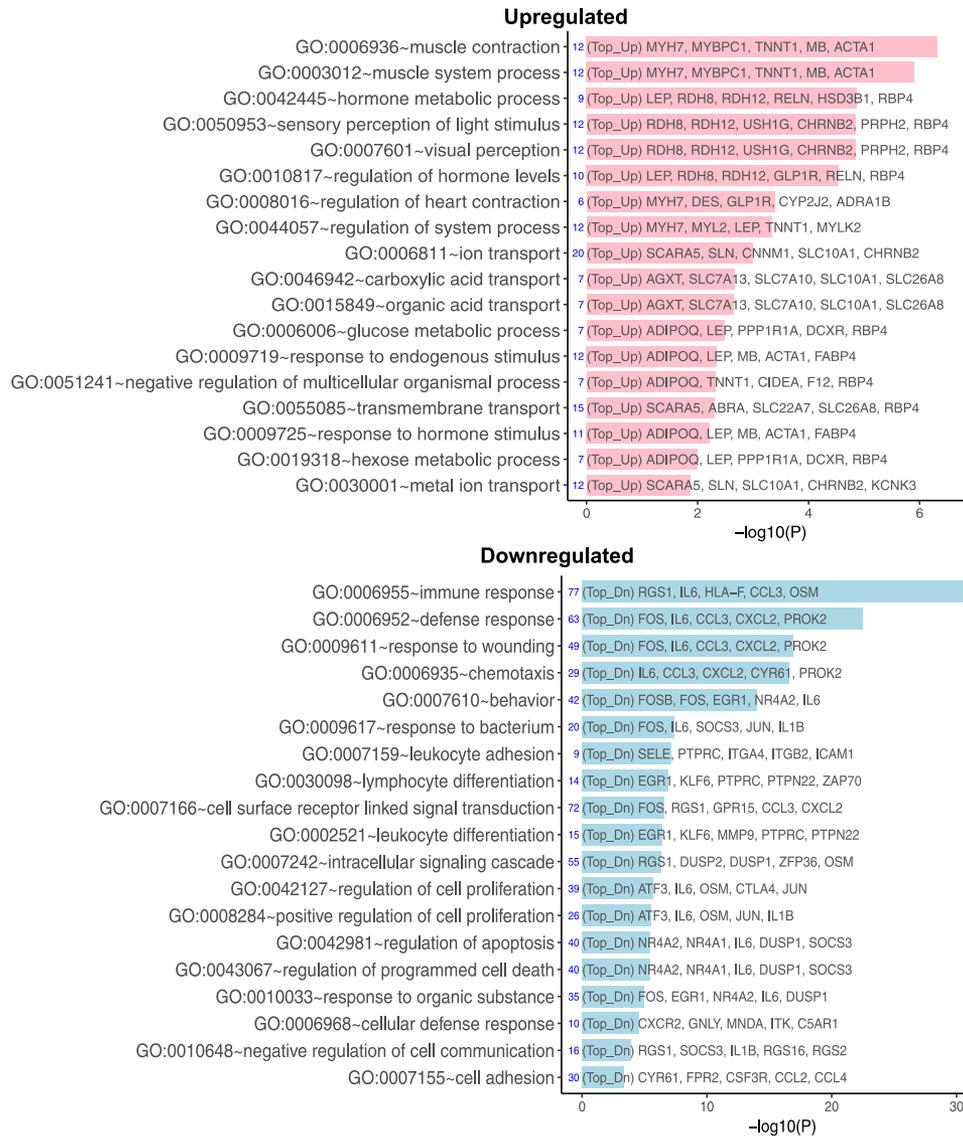
Figure 2—GO pathway analysis for DEGs for early DN vs. control. The top 20 enriched GO functions for upregulated (Top_Up, pink) and downregulated (Top_Dn, blue) DEGs are shown. $-\log10(P)$, $-\log_{10} P$ value.

Comparison of RNA-Seq Data of Kidney Biopsy Samples Between Advanced and Early DN

We next examined the gene expression between early DN and advanced DN. Figure 3 shows a heatmap and volcano plot that indicate the top 50 dysregulated genes in advanced DN versus early DN (a complete list of DEGs is included in Supplementary Excel File 1). Consistent with previous findings, many matrix- and inflammation-related genes were upregulated in advanced DN. GO analysis indicated that the top GO terms of upregulated genes are related to the immune response, whereas the downregulated DEGs are largely of transport and metabolic processes (Fig. 4). The pathway enrichment analysis for all DEGs showed that changes in the pathways were related to matrisome, fibrosis, inflammation, and metabolism (Supplementary Fig. 3). These findings are consistent with the high level of

inflammation, tubular dysfunction, and metabolic dysregulation observed in the kidneys of patients with advanced DN.

Comparison of RNA-Seq Data Among Control, Early, and Advanced DN Biopsy Samples

We then compared the gene expression in all three groups. Only a moderate number of genes changed in their expression progressively from control to early DN to advanced DN (five increased and seven decreased genes) (Fig. 5A and B and Supplementary Excel File 1). The analysis further identified 148 genes that increased in early DN versus control but decreased in advanced DN versus early DN (Fig. 5C and Supplementary Excel File 1). Interestingly, many of these genes were of the retinoic acid pathway, such as *RDH8*, *RDH12*, and *RBP4*; furthermore, *GLP1R* was among these genes. These data suggest
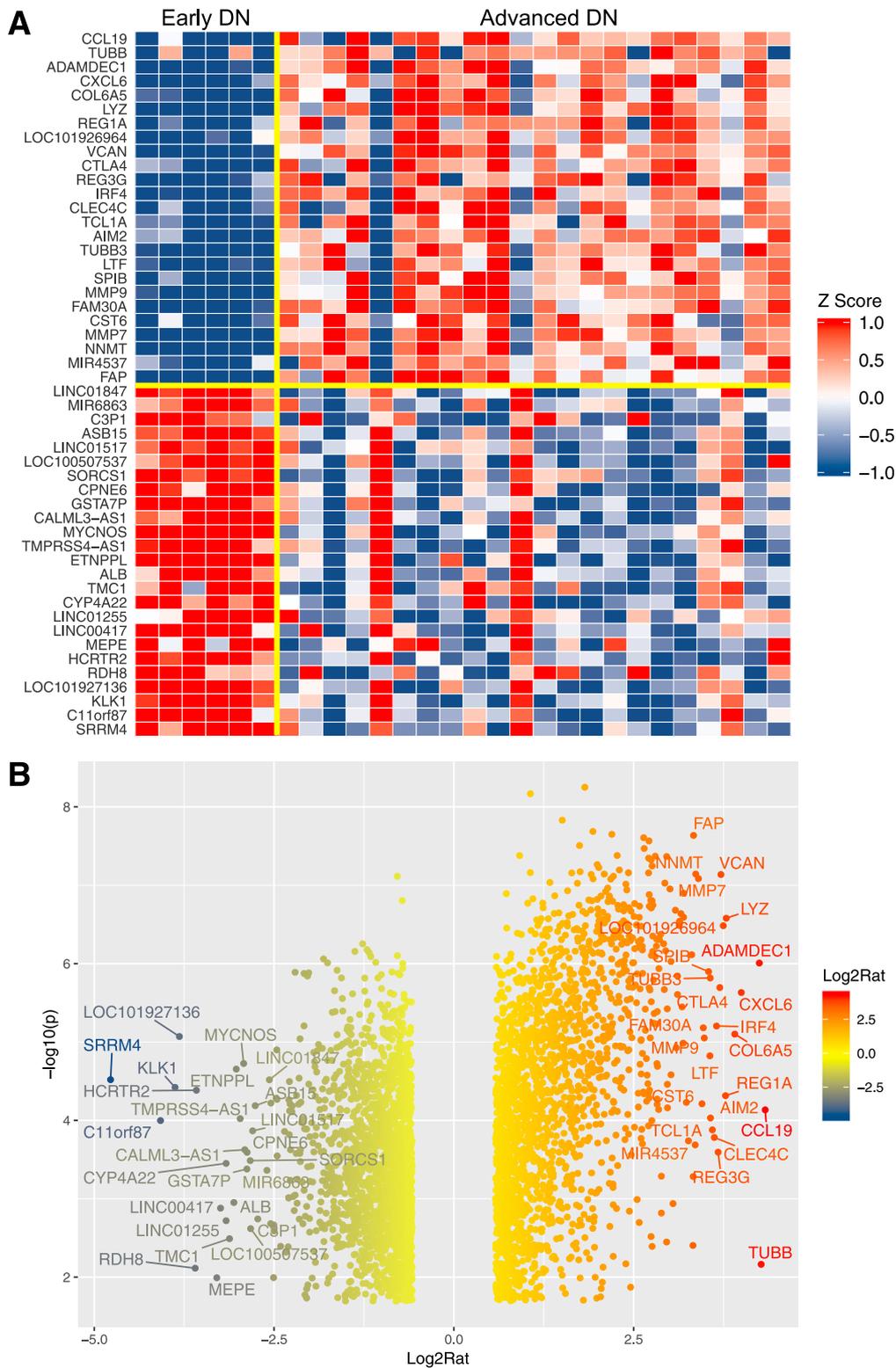
Figure 3—Differential gene expression analysis for advanced DN vs. early DN. *A*: Heatmap of the top 50 dysregulated genes (25 upregulated genes and 25 downregulated genes) in advanced DN samples. Data are *z* normalized for heatmap visualization. Each column represents an individual sample from the early DN or the advanced DN group. *B*: Volcano plot of dysregulated genes at limma FDR <0.05 and fold change >1.5 or less than −1.5. The top 50 DEGs are labeled on the plot. Log2Rat, $\log_2$ ratio; −log10(p), $-\log_{10} P$ value.

that retinoic acid and GLP1R agonists might have protective effects in early DN, thus preventing patients with diabetes from the progression of DN. The top GO terms

of the 148 genes are shown in Fig. 5*D* and a full list is found in Supplementary Excel File 1. We also identified 270 genes that were downregulated in patients with early

Figure 4—GO pathway analysis for DEGs for advanced DN vs. early DN. Top 20 enriched GO functions for upregulated (Top_Up, pink) and downregulated (Top_Dn, blue) DEGs are shown. −log10(P), −$\log_{10}$ P value.

DN versus control subjects but were significantly increased in advanced DN versus early DN (Fig. 5E and Supplementary Excel File 1). These genes were mostly related to immune response (Fig. 5F and Supplementary Excel File 1), suggesting that these genes may be involved in the promotion of DN through enhanced inflammation.
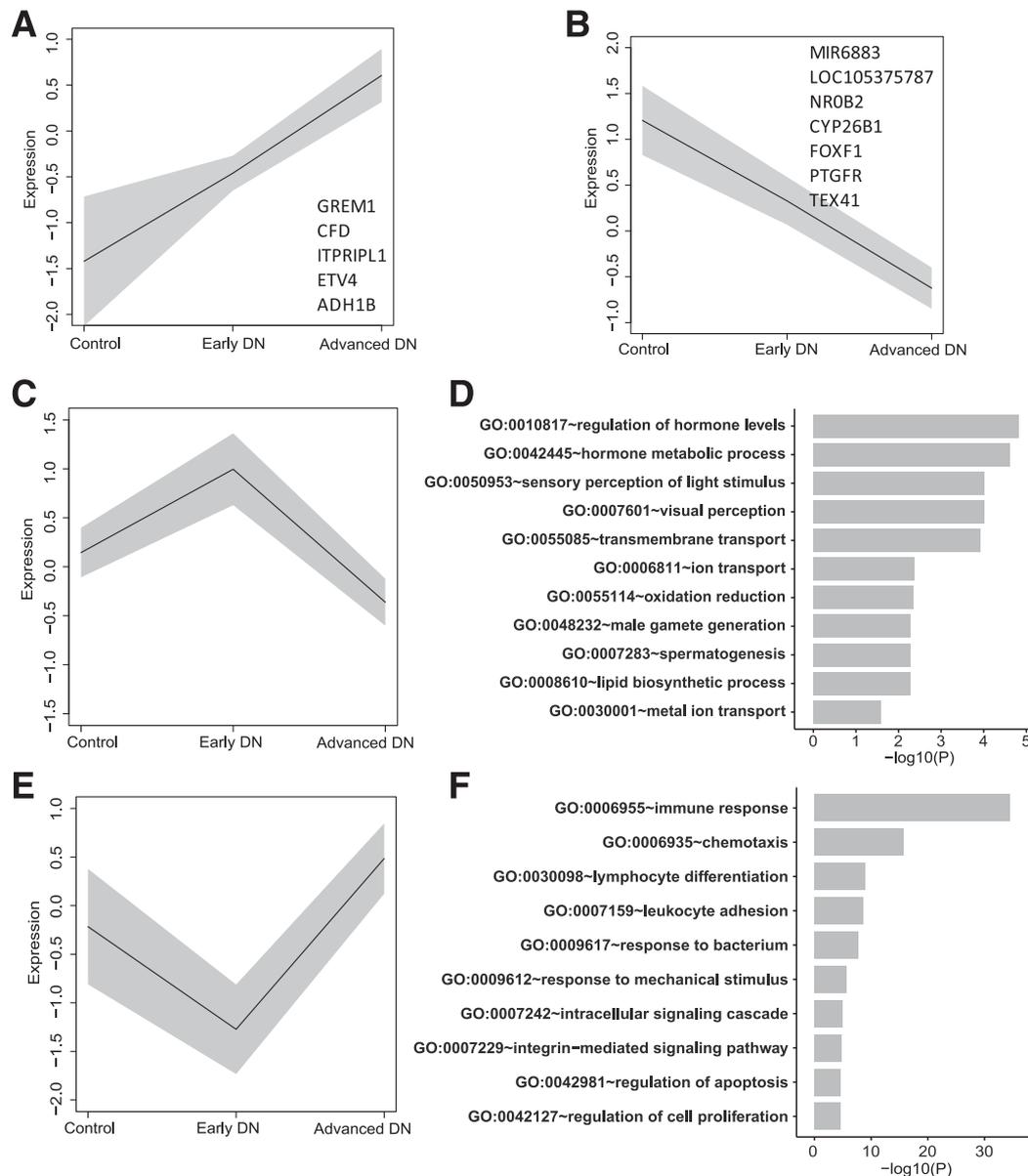
Figure 5—Gene expression change from control to DN states. Mean expression (black line) of genes and the SD of the mean (gray band) are shown for the control, early DN, and advanced DN groups. *A* and *B*: Mean expression of five upregulated (*A*) and seven downregulated (*B*) genes. *C*: Mean expression of 148 genes that were upregulated in early DN but downregulated in advanced DN. *D*: The top 10 GO enrichment terms for 148 DEGs in panel *C*. *E*: Mean expression of 270 genes that were downregulated in early DN but upregulated in advanced DN. *F*: The top 10 GO enrichment terms for 270 DEGs in panel *E*. $-\log10(P)$, $-\log_{10} P$ value.

Association of Genes With eGFR in Patients With DN

Additionally, we examined the association of genes in DN with eGFR in all patients with diabetes. GO terms related to iron transport and cell differentiation were positively associated with eGFR, while the immune response was at the top of the list of GO terms that were negatively associated with eGFR (Fig. 6*A*). Pathway enrichment analysis of combined DEGs showed phagocytosis regulation as one of the top pathways associated with eGFR (Supplementary Fig. 4). Small G-protein regulation (RhoA and Rac1), fibrosis, and inflammatory pathways were also highly associated with eGFR. The individual list of genes

associated with eGFR is included in Supplementary Excel File 2.

Association of Genes With Kidney Histological Scores in All Patients With DN

We also examined the correlation between genes and histological scores in all patients with DN. Supplementary Table 2 shows the scoring of the individual histological parameters in patients with early or advanced DN. The most enriched GO pathways that correlated with histology scores in all DN samples are summarized in Fig. 6*B*. Segmental glomerulosclerosis (GS), tubular atrophy, and
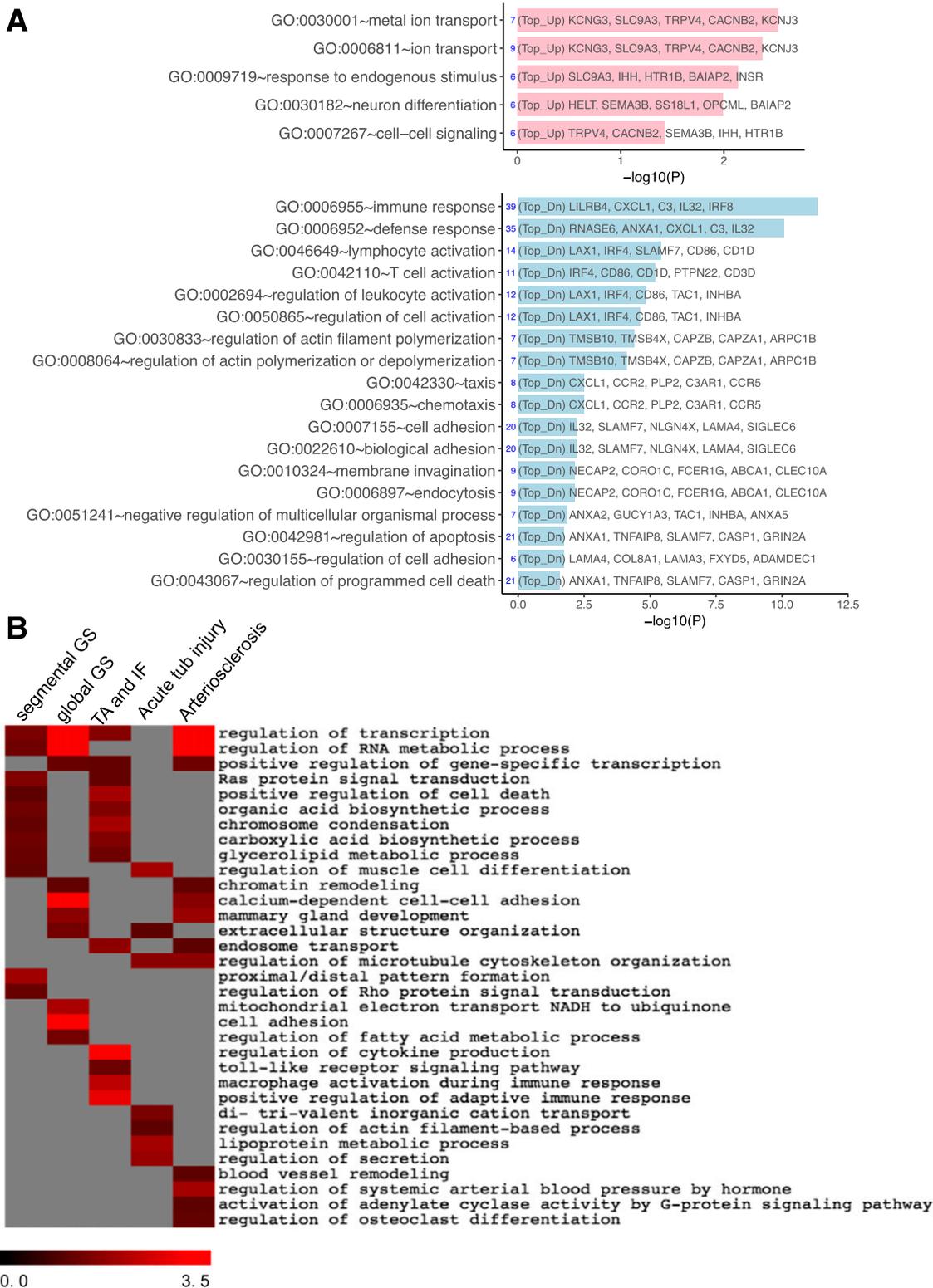
A

B

Figure 6—Top enriched GO functions for eGFR and histological parameters in all DN samples. *A*: The bar chart shows the top enriched functions for DEGs that correlated positively (pink) or negatively (blue) with eGFR. *B*: A summary heatmap of the top enriched functions shows pathways that are correlated with each histological parameter. Scale bar indicates significance (shown as $-\log_{10} P$ value [$-\log_{10}(P)$]). IF, interstitial fibrosis; TA, tubular atrophy; Top_Dn, top downregulated DEGs; Top_Up, top upregulated DEGs; tub, tubular.

fibrosis had several shared GO pathways, such as Ras protein signaling, cell death, and several metabolic pathways. Global GS and arteriosclerosis also shared several

pathways, such as chromatin remodeling and calcium-dependent cell-cell adhesion. Regulation of microtubule cytoskeleton was a common pathway between acute tubular

injury and arteriosclerosis. However, each histological parameter also had distinct GO terms. For example, segmental GS had a unique feature on Rho protein signaling, which may be related to podocyte dysfunction. Mitochondrial dysfunction and fatty acid disturbance were associated more with global GS. Multiple inflammation-related pathways were associated with tubular atrophy/interstitial fibrosis. However, the acute tubular injury was associated with actin filament, iron transport, and secretion. Arteriosclerosis had specific GO terms on blood vessel remodeling, hormonal regulation of systemic arterial blood pressure, activation of adenylate cycle activity, and regulation of osteoclast differentiation. The latter might be related to the vascular calcification. Overall, these pathway analyses highlight the potential underlying molecular mechanisms for the different pathological changes observed in DN. The full list of DEGs associated with each histological parameter is included in Supplementary Excel File 3.

Deconvolution of RNA-Seq Data for the Representation of Kidney Cell Types

Because the biopsied tissue samples used in this study contain a heterogeneous mixture of kidney cell types, we used an R-based algorithm, CIBERSORT (11), to deconvolve the current data set to estimate the relative fractions of diverse kidney cell types. For this, we used the recently published scRNA-seq data set from human kidneys as a reference from Wu et al. (12). Enumeration of the data showed distinct clusters of kidney cells in control, early DN, and advanced DN kidneys (Fig. 7A). We found a significant reduction of proximal tubular and collecting cells in advanced DN compared with early DN and control samples. Interestingly, endothelial cells appeared to be decreased in both early and advanced DN compared with control, suggesting that kidney (glomerular and peritubular) endothelial cell injury may occur in early DN (Fig. 7A). There was an increase of monocytes, fibroblasts, myofibroblasts, B cells, and plasma cells in advanced DN compared with early DN and control samples. Since the scRNA-seq data used as a reference by Wu et al. did not include macrophages as a cell cluster, we could not determine whether there was a change in the macrophage population in DN. Therefore, we used the scRNA-seq data from normal mouse kidneys from Park et al. (13) as a reference, which showed a significant increase of macrophages and a small but significant increase in both T and B cells in advanced DN compared with early DN and control (Fig. 7B). These analyses are consistent with previous histological observations of increased inflammation, fibrosis, and tubular cell loss in the kidneys of patients with advanced DN compared with those from patients with early DN and control patients.

The above deconvolution analysis showed a significant reduction of tubular cells in advanced DN. However, we could not appreciate the changes of podocytes because of the small portion of podocytes in the kidney cortices and

the limitations of the analysis. Therefore, we compared the expression of both glomerular and tubular cell–specific markers and found that there was a reduction of both podocyte- and tubular cell–specific markers in advanced DN. The full list of kidney cell–specific DEGs is presented in Supplementary Excel File 4.

Since the changes of kidney cell population may affect the data analysis, we rechecked the DEGs and the GO pathways from early DN versus control and advanced DN to early DN after adjusting for the differences in cell population, as estimated from the deconvolution analysis. The adjusted analysis (Supplementary Figs. 5 and 6) largely showed similar findings to the unadjusted analysis (Figs. 2 and 4). In addition, we compared the DEGs from this study with the previously published transcriptomic data sets from Pan et al. (7) (GSE96804), a microarray analysis of glomerular transcriptome in DN, and Woroniecka et al. (5) (GSE30122), a microarray analysis of glomerular and tubular transcriptomes. As shown in Supplementary Table 3, there was an ~30% overlap in DEGs and >60% overlap in most of the enriched pathways compared with both sets. Given that these published data sets are of glomerular or tubular compartment–specific microarray analyses and that our data set is of RNA-seq data of biopsied whole-cortex samples, the overlap is quite significant between our and these two studies.

Validation of the Findings From Our RNA-Seq Data

We next confirmed by immunohistochemical analysis the change in expression of several of the DEGs identified. We were particularly interested in the genes that were increased in early DN but decreased in advanced DN because they may represent genes that may have a protective role against the progression of DN. Among these, we selected RBP4 and GLP1R for further validation by immunostaining in the kidney sections of patients with DN because their expression has not been well characterized in diabetic kidneys. As shown in Fig. 8A, we confirmed that RBP4 and GLP1R increased in early DN but decreased in advanced DN. RBP4 staining localized mostly in renal tubular cells in early DN, which largely colocalized with proximal tubule marker AQP1 (Supplementary Fig. 7), while the staining was very weak in the kidneys with advanced DN (Fig. 8A and Supplementary Fig. 8A). GLP1R localized in both glomerular and proximal tubular cells in early DN kidneys and colocalized with AQP1 in the tubular compartment (Supplementary Fig. 7), but the staining was also very weak in advanced DN (Fig. 8A and Supplementary Fig. 8A). For genes that are increased in advanced DN but not in early DN, we selected two immune response genes for validation by immunostaining. We found that expression of both interleukin 6 (IL-6) and IL-1B was significantly increased in advanced DN kidneys but not in early DN kidneys (Fig. 8B and Supplementary Fig. 8B).

We also performed immunostaining for the markers of immune cells to validate the findings from the deconvolution data. We found that the staining for M2-macrophage
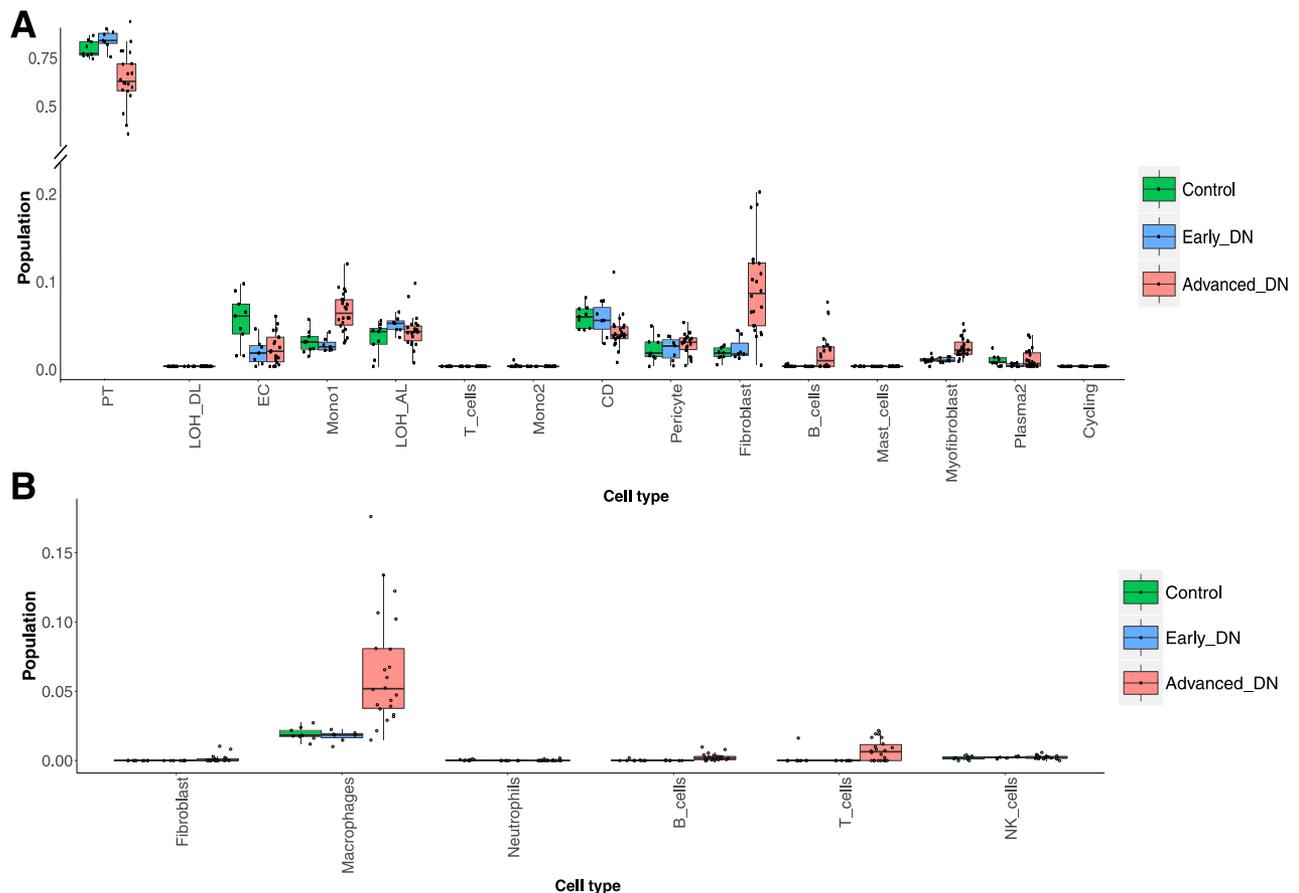
Figure 7—Estimates of cell components by deconvolution analysis in control, early, and advanced DN samples. *A*: Box plots of deconvoluted cell population across all cell types estimated by Wu et al. (12). *B*: Box plots of deconvoluted cell population across immune cells estimated by Park et al. (13). CD, collecting duct; EC, endothelial cell; LOH_AL, loop of Henle, ascending limb; LOH_DL, loop of Henle, distal limb; Mono1, monocyte type 1; Mono2, monocyte type 2; NK, natural killer; PT, proximal tubule.

marker MRC1 increased significantly in advanced DN compared with early DN (Fig. 8*C*). There was also a significant increase of staining for T- and B-cell markers CD3 and CD20, respectively, in advanced DN compared with early DN kidneys (Fig. 8*C* and Supplementary Fig. 8*C*). These data are consistent with our deconvolution analysis and support an important role of these immune cells in the progression of DN.

DISCUSSION

In the current study, we performed transcriptomic studies of whole-kidney biopsy samples from patients with either early or advanced DN compared with control nephrectomy samples from patients without diabetes obtained from Shanghai Jiao Tong University Affiliated Sixth People's Hospital. For this study, we chose to use the whole-kidney biopsy samples for analysis. The rationale for this was mainly because of the limited amount of tissue that is typically available in biopsy samples and, importantly, because the dissection and digestion processes could alter the transcriptomic profiles by induction of stress-related gene expression as described in a previous study (15). The digestion process in particular may also have differential

effects between nephrectomy and native kidney biopsy samples as well as between normal and diseased kidneys, thereby creating further artificial differences in gene expression between experimental groups. However, the limitation to our approach is that since kidney cortices contain mostly a tubulointerstitial compartment, the transcriptomic data will reflect mostly the mRNA expression of cells in this compartment. Indeed, the largest population of cell types identified by deconvolution analysis was proximal tubules. While the isolation of glomeruli would yield more specific information on glomerular cells, the data will nevertheless represent those of heterogeneous glomerular cell types. To circumvent this issue, we had performed isolation of specific glomerular cells from fluorescently labeled cells in mice (21,22), which cannot be done in human tissue samples. More recently, scRNA-seq has emerged as a powerful tool for observing gene expression at a single-cell level (23,24). However, the depth and number of the genes detected by scRNA-seq are still limited compared with the bulk RNA-seq (25). As mentioned above, the scRNA-seq data can be significantly affected by the digestion process because the procedure requires more stringent digestion of kidney tissues into the single cells.
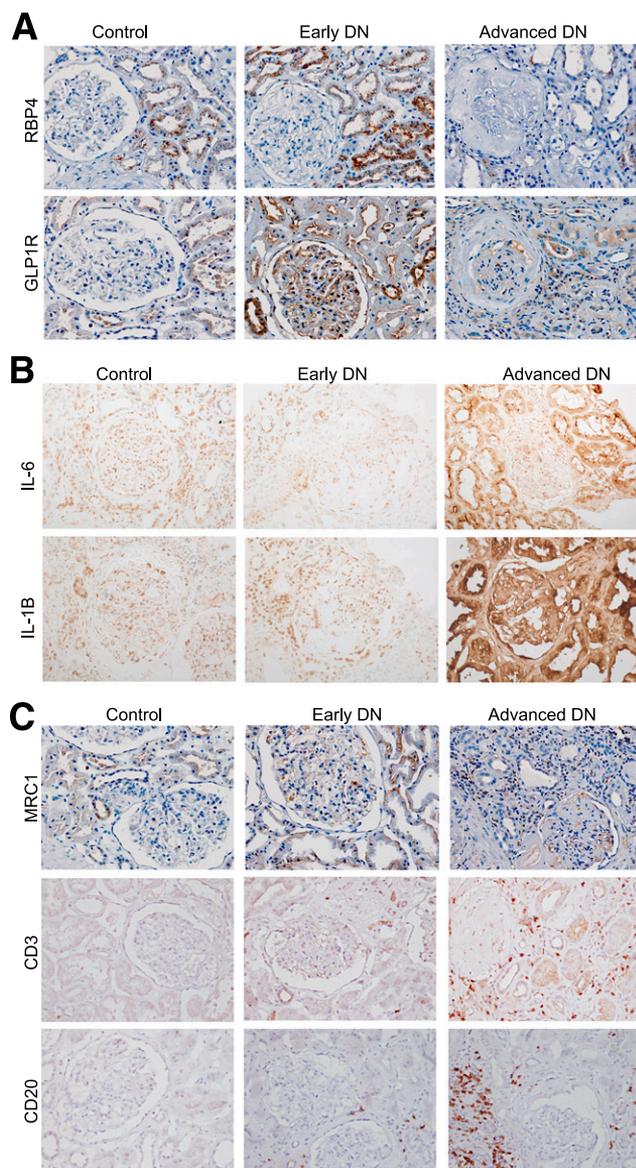
Figure 8—Immunostaining of genes that are altered in DN. *A*: Representative images of RBP4 and GLP1R immunostaining in control and DN kidneys. *B*: Representative images of IL-6 and IL-1B immunostaining in control and DN kidneys. *C*: Representative images of CD20, CD3, and MRC1 immunostaining in control and DN kidneys. Semiquantification of immunostaining is shown in Supplementary Fig. 8.

In addition, digestion of the human kidney biopsy samples for scRNA-seq is challenging to perform because of the limited amount of available material (26,27). Therefore, each technical approach has distinct advantages and disadvantages but together can provide complementary information to better understand the pathogenesis of DN.

There are a few advantages of the transcriptomic analysis described in this study. First, we used the RNA-seq approach, rather than microarray analyses used in most previous studies, to study the transcriptome of diabetic kidneys. The limitation of microarray analysis is that not all genes can be detected with currently available chips. In

addition, our study provides data on the changes of non-coding RNAs in the diabetic kidney. For example, we found several noncoding RNAs among the DEGs between early and advanced DN that included LOC101926964, LOC101927136, LOC100507537, LINC00417, and LINC01255. We found that miR-3189 was upregulated in early DN compared with control, and it is known that miR-3189 is a potent regulator of cell apoptosis through the p53-dependent pathway (28). Therefore, we will confirm whether mature miR-3189 is indeed upregulated in patients with early DN compared with control patients in future investigations.

Another strength of our study is that we were able to include several samples from patients with early DN in addition to samples from patients with advanced DN. Although we only had samples from six patients with clinical and biopsy-proven early DN, the transcriptomic data from these patients provide useful insight into the early disease process of DN. These six patients have the same duration of diabetes and other clinical parameters as those with advanced DN and, therefore, are likely to be nonprogressors. To further support this, we have followed these six patients for 2–3 years since their enrollment and found that they had stable renal function and UACR. Previous studies included only patients with advanced DN (4,5), and the only other kidney transcriptome of early DN was performed in the Pima Indian study (6). Because our study is of patients of Asian origin, it would be informative to compare the transcriptomic data of patients with DN from different ethnic backgrounds.

Also, by using the recently published scRNA-seq data sets from normal human and mouse kidneys, we were able to deconvolve the bulk RNA-seq data from DN samples to estimate the changes occurring in specific kidney cell types in DN. Because of the variations of methodology and analysis among the currently available scRNA-seq data sets, we used two scRNA-seq data sets as a reference: human (26) and mouse (13) whole-kidney cortices. Even though the deconvolution data generated from the two data sets had some differences, the overall results suggest that there is an increase of monocytes/macrophages, fibroblasts, and myofibroblasts in DN kidneys. We further confirmed this by immunostaining of the kidney sections from the same patients. Thus, our data further support the critical role of macrophages in the pathogenesis of human DN. We are aware of the limitations of this analysis, but we believe that this approach will be significantly improved when we have more reliable scRNA-seq data from patients with DN.

When we examined the DEGs between early and advanced DN and control, we identified a group of genes that was highly expressed in early DN but suppressed in advanced DN, such as genes in the retinoic acid pathway and GLP1R. These likely represent a group of genes that have renoprotective effects in the early stage of DN, consistent with the nonprogressive nature of these patients. In addition, we were able to validate the

expression of RBP4 and GLP1R in the kidney tissues of these patients. The role of RBP4 is to deliver retinol to the target tissue (29) and has been shown to be associated with renal function in patients with diabetes (30). Retinoic acid has been shown to have protective effects in vitro and in animal models of kidney disease (19,20). Recent studies suggested that local synthesis of retinoic acid appears impaired in the diseased kidney and contributes to the progression of kidney disease (31). Interestingly, genetic variations of RBP4 are associated with early DN but not advanced DN (32). Our study suggests that local expression of RBP4 might be elevated in early DN and contribute to local retinoic acid synthesis and protects patients from the progression of DN. Future studies are required to further examine the role of RBP4 in the progression of DN. GLP1R agonists have been shown to have renal protective effects in animal models and humans with DN (33–35). Our study suggests that GLP1R expression was increased locally in the kidney to protect the kidney from injury in early DN. Immunostaining of GLP1R indicated that its expression is indeed increased in both glomerular and tubular cells in early DN. How GLP1R mediates the effects of its agonists in local kidney cells requires further studies.

Interestingly, we found that expression of the immune response or inflammatory genes were suppressed in early DN but highly upregulated in advanced DN. The suppression of immune response genes in early DN could be partially due to mildly elevated inflammatory status of control nephrectomy samples. But more likely, these patients with early DN are nonprogressors and resistant to the progression of DN, and therefore, the inflammation is suppressed. The marked increase in immune response and inflammatory genes in patients with advanced DN confirms a critical role of inflammation in the progression of DN and is consistent with the previous transcriptomic data showing that the Janus kinase-STAT and nuclear factor-κB pathways are highly enriched in diabetic kidneys (2,4) and with the findings in Woroniecka et al. (5). Consistently, our deconvolution data suggest an increase in immune cells in the diabetic kidney, such as macrophages. Together, these data support a critical role of inflammation in the progression of DN.

We also performed correlation studies of gene transcripts with renal function (eGFR). Interestingly, DEGs related to phagocytosis and inflammatory pathways were negatively correlated with eGFR, suggesting again the role of macrophages and immune response in the progression of DN. DEGs related to iron transports were positively correlated with eGFR, indicating that tubular cell injury (loss of iron transports) contributes to the progression of DN.

DN has classic pathological features, with injury in the glomeruli, tubule/interstitium, and blood vessels (14,36). Different patients may present more injury in one of these compartments than others. However, the underlying molecular mechanisms of each pathological feature in DN

remain unclear. Therefore, we studied the association of renal transcripts with the histological scores of individual pathological features obtained from patients with both early and advanced DN. Interestingly, we found that the segmental GS, global GS, and arteriosclerosis share some but have their own unique GO terms. Also, tubular fibrosis score was highly associated with the DEGs related to the immune response and inflammatory pathways. Since tubular fibrosis is known to be tightly associated with eGFR, our data show that the DEGs related to inflammation pathways correlate with eGFR. Interestingly, we also found several specific GO terms associated with arteriosclerosis, such as blood vessel remodeling and hormonal regulation of blood pressure. Several genes involved in bone metabolism were also associated with arteriosclerosis, and they may be involved in vascular calcification in patients with chronic kidney disease. A similar study was performed recently in a Pima Indian cohort with DN (6). The authors reported that the cortical interstitial fractional volume, an index of tubulointerstitial damage, correlated significantly with the transcripts enriched for pathways associated with mitochondrial dysfunction, inflammation, migratory mechanisms, and tubular metabolic functions. Further studies are required to determine how these DEGs contribute to the specific pathological changes observed in DN.

In conclusion, our study provides the transcriptomic data of patients with early and advanced DN compared with normal tissues from control patients from nephrectomy samples in a Chinese population of patients with diabetes. The correlation of renal transcripts with renal function and pathological changes will help us to further understand the underlying molecular mechanisms contributing to the progression of DN. We believe that whole-kidney transcriptomic data and scRNA-seq data will be complementary, and future sophisticated computational programs could help to better dissect the mechanisms of individual kidney cell injury by combining these two data sets. Finally, we believe that the data generated here could be an important resource for the renal community to further dissect the pathogenesis of DN.

References

1. Collins AJ, Foley RN, Chavers B, et al. US Renal Data System 2013 Annual Data Report. Am J Kidney Dis 2014;63(Suppl.):A7

2. Schmid H, Boucherot A, Yasuda Y, et al.; European Renal cDNA Bank (ERCB) Consortium. Modular activation of nuclear factor-kappaB transcriptional programs in human diabetic nephropathy. Diabetes 2006;55:2993–3003

3. Lindenmeyer MT, Kretzler M, Boucherot A, et al. Interstitial vascular rarefaction and reduced VEGF-A expression in human diabetic nephropathy. J Am Soc Nephrol 2007;18:1765–1776

4. Berthier CC, Zhang H, Schin M, et al. Enhanced expression of Janus kinase-signal transducer and activator of transcription pathway members in human diabetic nephropathy. Diabetes 2009;58:469–477

5. Woroniecka KI, Park AS, Mohtat D, Thomas DB, Pullman JM, Susztak K. Transcriptome analysis of human diabetic kidney disease. Diabetes 2011;60:2354–2369

6. Nair V, Komorowsky CV, Weil EJ, et al. A molecular morphometric approach to diabetic kidney disease can link structure to function and outcome. Kidney Int 2018;93:439–449

7. Pan Y, Jiang S, Hou Q, et al. Dissection of glomerular transcriptional profile in patients with diabetic nephropathy: SRGAP2a protects podocyte structure and function. Diabetes 2018;67:717–730

8. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013;29:15–21

9. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015;43:e47

10. Huang W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009;4:44–57

11. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods 2015;12:453–457

12. Wu H, Malone AF, Donnelly EL, et al. Single-cell transcriptomics of a human kidney allograft biopsy specimen defines a diverse inflammatory response. J Am Soc Nephrol 2018;29:2069–2080

13. Park J, Shrestha R, Qiu C, et al. Single-cell transcriptomics of the mouse kidney reveals potential cellular targets of kidney disease. Science 2018;360:758–763

14. Tervaert TW, Mooyaart AL, Amann K, et al.; Renal Pathology Society. Pathologic classification of diabetic nephropathy. J Am Soc Nephrol 2010;21:556–563

15. van den Brink SC, Sage F, Vértesy Á, et al. Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. Nat Methods 2017;14:935–936

16. Brennan EP, Mohan M, McClelland A, et al. Lipoxins regulate the early growth response-1 network and reverse diabetic kidney disease. J Am Soc Nephrol 2018;29:1437–1448

17. Feigerlová E, Battaglia-Hsu SF. IL-6 signaling in diabetic nephropathy: from pathophysiology to therapeutic perspectives. Cytokine Growth Factor Rev 2017;37:57–65

18. Chung AC, Lan HY. Chemokines in renal injury. J Am Soc Nephrol 2011;22:802–809

19. He JC, Lu TC, Fleet M, et al. Retinoic acid inhibits HIV-1-induced podocyte proliferation through the cAMP pathway. J Am Soc Nephrol 2007;18:93–102

20. Ratnam KK, Feng X, Chuang PY, et al. Role of the retinoic acid receptor-α in HIV-associated nephropathy. Kidney Int 2011;79:624–634

21. Fu J, Wei C, Lee K, et al. Comparison of glomerular and podocyte mRNA profiles in streptozotocin-induced diabetes. J Am Soc Nephrol 2016;27:1006–1014

22. Fu J, Wei C, Zhang W, et al. Gene expression profiles of glomerular endothelial cells support their role in the glomerulopathy of diabetic mice. Kidney Int 2018;94:326–345

23. Wang Y, Navin NE. Advances and applications of single-cell sequencing technologies. Mol Cell 2015;58:598–609

24. Gawad C, Koh W, Quake SR. Single-cell genome sequencing: current state of the science. Nat Rev Genet 2016;17:175–188

25. Wu H, Humphreys BD. The promise of single-cell RNA sequencing for kidney disease investigation. Kidney Int 2017;92:1334–1342

26. Wu H, Kirita Y, Donnelly EL, Humphreys BD. Advantages of single-nucleus over single-cell RNA sequencing of adult kidney: rare cell types and novel cell states revealed in fibrosis. J Am Soc Nephrol 2019;30:23–32

27. Der E, Ranabothu S, Suryawanshi H, et al. Single cell RNA sequencing to dissect the molecular heterogeneity in lupus nephritis. JCI Insight 2017;2:e93009

28. Jones MF, Li XL, Subramanian M, et al. Growth differentiation factor-15 encodes a novel microRNA 3189 that functions as a potent regulator of cell death. Cell Death Differ 2015;22:1641–1653

29. Blaner WS. Retinol-binding protein: the serum transport protein for vitamin A. Endocr Rev 1989;10:308–316

30. Henze A, Frey SK, Raila J, et al. Evidence that kidney function but not type 2 diabetes determines retinol-binding protein 4 serum levels. Diabetes 2008;57:3323–3326

31. Li X, Dai Y, Chuang PY, He JC. Induction of retinol dehydrogenase 9 expression in podocytes attenuates kidney injury. J Am Soc Nephrol 2014;25:1933–1941

32. Ng DP, Salim A, Lim XL, Nurbaya S. Estimated glomerular filtration rate and its association with the retinol-binding protein 4 (RBP4) locus on human chromosome 10q23. Nephrol Dial Transplant 2012;27:1511–1515

33. Imamura S, Hirai K, Hirai A. The glucagon-like peptide-1 receptor agonist, liraglutide, attenuates the progression of overt diabetic nephropathy in type 2 diabetic patients. Tohoku J Exp Med 2013;231:57–61

34. DeFronzo RA. Combination therapy with GLP-1 receptor agonist and SGLT2 inhibitor. Diabetes Obes Metab 2017;19:1353–1362

35. Fujita H, Morii T, Fujishima H, et al. The protective roles of GLP-1R signaling in diabetic nephropathy: possible mechanism and therapeutic potential. Kidney Int 2014;85:579–589

36. Haneda M, Utsunomiya K, Koya D, et al.; Joint Committee on Diabetic Nephropathy. A new Classification of Diabetic Nephropathy 2014: a report from Joint Committee on Diabetic Nephropathy. J Diabetes Investig 2015;6:242–246