Molecular Profiling of Diabetic Mouse Kidney Reveals Novel Genes Linked to Glomerular Disease

Katalin Susztak,[1] Erwin Böttinger,[1,2] Akiva Novetsky,[1] Dan Liang,[1] Yanqing Zhu,[3] Emilio Ciccone,[3] Dona Wu,[1,2] Stephen Dunn,[3] Peter McCue,[4] Kumar Sharma[3]

To describe gene expression changes that characterize the development of diabetic nephropathy, we performed microarray and phenotype analysis on kidneys from *db/db* mice (a model of type 2 diabetes), streptozotocin-induced diabetic C57BL/6J mice (a model of type 1 diabetes), and nondiabetic controls. Statistical comparisons were implemented based on phenotypic outcome characteristics of the animals. We used weighted vote–based supervised analytical methods to find genes whose expression can classify samples based on the presence or absence of mesangial matrix expansion, the best indicator for the development of end-stage renal disease in humans. We identified hydroxysteroid dehydrogenase-3β isotype 4 and osteopontin as lead classifier genes in relation to the mesangial matrix expansion phenotype. We used the expression levels of these genes in the kidney to classify a separate group of animals for the absence or presence of diabetic glomerulopathy with a high degree of precision. Immunohistochemical analysis of murine and human diabetic kidney samples showed that both markers were expressed in podocytes in the glomeruli and followed regulation similar to that observed in the microarray. The application of phenotype-based statistical modeling approaches has led to the identification of new markers for the development of diabetic kidney disease. *Diabetes* 53:784–794, 2004

From the [1]Department of Medicine, Division of Nephrology, Albert Einstein College of Medicine, Bronx, New York; the [2]Department of Molecular Genetics, Albert Einstein College of Medicine, Bronx, New York; the [3]Department of Medicine, Division of Nephrology, Dorrance Hamilton Research Laboratories, Thomas Jefferson University, Philadelphia, Pennsylvania; and the [4]Department of Anatomy, Cell Biology and Pathology, Division of Nephrology, Dorrance Hamilton Research Laboratories, Thomas Jefferson University, Philadelphia, Pennsylvania.

Address correspondence and reprint requests to Kumar Sharma, MD, Associate Professor of Medicine, Dorrance Hamilton Research Laboratories, Division of Nephrology, Department of Medicine, Thomas Jefferson University, Suite 353, Jefferson Alumni Hall, 1020 Locust St., Philadelphia, PA 19107. E-mail: kumar.sharma@jefferson.edu. Or to Erwin Bottinger, Professor of Medicine, Mount Sinai Medical Center, One Gustave L. Levy Place, Box 1118, New York, NY 10029. E-mail: erwin.bottinger@mssm.edu.

Received for publication 29 October 2003 and accepted in revised form 15 December 2003.

Additional information for this article can be found in an online appendix at http://diabetes.diabetesjournals.org.

K.S. and E.B. contributed equally to this work.

CBI, channel background intensity; CHI, channel signal intensity; FDR, false discovery rate; HSD3β4, hydroxysteroid dehydrogenase-3β isotype 4; KAP, kidney androgen–regulated protein; NSI, normal signal intensity; OPN, osteopontin; QRT-PCR, quantitative RT-PCR; ROC, receiver operator curve; SAM, Statistical Analysis of Microarray Data; STZ, streptozotocin; TGF-β, transforming growth factor-β.

As diabetic nephropathy develops in as many as 30% of diabetic patients (1), the identification of genes that are involved in predicting diabetic nephropathy and play a role in its pathogenesis is a topic of great importance. Approaches using candidate gene analysis have improved our understanding of key pathways in addressing some aspects related to pathogenesis (2); however, polymorphisms of many candidate genes have not been established to confer genetic susceptibility in large epidemiological studies (3). Microarray analysis is a novel tool that may help to establish new diagnostic and predictive molecular targets for accurate assessment of risk for diabetic nephropathy in diabetic individuals without renal manifestations. In addition, whole genome analysis via microarrays can identify new genes and pathways that are important in the pathophysiology of diabetic nephropathy (4).

Common features of both type 1 and type 2 diabetic nephropathy are the development of albuminuria and later diffuse mesangial matrix expansion. While albuminuria can regress in up to 50% of patients (5), the extent of mesangial matrix expansion correlates closely with progression of nephropathy (6); therefore, genes that can predict the development of mesangial matrix expansion are highly relevant. Recent reports have identified genes by microarray screens whose expression in kidneys of diabetic animals is either increased or decreased compared with controls (7,8). However, it is often difficult to select individual genes from microarray datasets that are associated with (and can therefore diagnose) a single phenotypic outcome. An alternative approach could be using kidney-specific cell lines; however, this approach could miss important cell-cell interactions in the development of diabetic nephropathy (9,10).

In this study, we used several approaches to analyze a large microarray dataset in order to find genes with a strong association and high diagnostic value for the development of diabetic glomerulosclerosis in mice. We first used different models and strains of diabetic mice representing a spectrum of diabetic nephropathy. We next defined experimental groups based on differences in the manifestation of characteristic phenotypes, such as absence or presence of hyperglycemia or mesangial matrix expansion. This design is likely to identify differential gene expression that is more closely related to pathomechanisms affecting the tissue of interest because other phenotypic and genotypic differences are not considered. Super-

vised statistical methods were applied to find genes whose expression can classify animals into separate groups based on phenotypic characteristics. This approach lead to the identification and validation of new marker genes with strong correlation for the development of diabetic nephropathy.

RESEARCH DESIGN AND METHODS

Male C57/B6 and 129SvJ, *db/db* (*Lepr$^{db/db}$*), and *db/m* mice (on C57BLKs/J background) were purchased from The Jackson Laboratories (Bar Harbor, ME). Type 1 diabetes was induced by low-dose streptozotocin (STZ) injection as detailed by the NIDDK Consortium for Animal Models of Diabetic Complications' (AMDCC) protocol (available from http://www.amdcc.org). The *db/db* mice were maintained either on normal mouse chow (Purina 5001; Purina, Quakertown, PA) or on high-protein diet (40% protein and 4% salt). All animal protocols and procedures were approved by the Institutional Animal Care and Use Committee at Thomas Jefferson University.

For determination of urinary albumin excretion, mice were placed into individual metabolic cages and urinary albumin was measured by enzyme-linked immunoassay specific for mouse albumin (Albuwell; Exocell, Philadelphia, PA). Blood glucose was measured using a colorimetric assay (Sigma). A pathologist who was unaware of the experimental protocol analyzed glomerular histology on periodic acid Schiff–stained sections. Mesangial matrix expansion was evaluated on 50 glomeruli per kidney using a score of 1 for minimal, 2 for mild mesangial matrix expansion, 3 for moderate, and 4 for diffuse mesangial matrix expansion.

Microarray procedures. Total kidney RNA was extracted using Trizol reagent (Invitrogen). RNA samples, without evidence of degradation, were used for microarray analysis. Common reference RNA standard was prepared from female and male C57B/6J mice in a 1:1 ratio, using a mixture of RNA from liver, kidney, and spleen at a 4:2:1 RNA weight ratio, respectively. Mouse cDNA arrays were obtained from the Albert Einstein College of Medicine cDNA Microarray Facility (available from http://www.aecom.yu.edu/home/molgen/facilities.html). Each slide contained an unbiased, random collection of 9,557 cDNA probe elements derived from the sequence-verified GEM1 clone set (Incyte Genomics, Palo Alto, CA).

Microarray procedures were performed as described previously with minor modifications (11) (available from http://biotech.aecom.yu.edu). In brief, RNA (100 μg) from a single experimental kidney was reverse transcribed and labeled with Cy5-fluorescent nucleotides (Amersham, Piscataway, NJ). Identical aliquots (100 μg) of the common reference RNA were labeled with Cy3-fluorescent nucleotides by reverse transcription. For each experimental animal, Cy5-labeled kidney cDNA was cohybridized with Cy3-labeled common reference cDNA. Hybridized slides were scanned using a GenePix 4000b scanner (Axon Instruments, Union City, CA), and raw data files were generated containing measurements of signal and background fluorescence emissions. The raw data files and the microarray platform for all of the hybridizations can be viewed at www.ncbi.nlm.nih.gov/geo with accession number GPL-409.

Quality control and data analysis. Data from spots containing artifacts were eliminated from further analysis. Data were also filtered, and only spots with average channel signal intensity (CHI) over the average channel background intensity (CHB) + 2 SD were further analyzed, while spots with average normal signal intensity (NSI) less than or equal to average CHB + 2 SD were flagged and eliminated from further analysis. The ratio of Cy5 (experimental sample) and Cy3 (common reference RNA) was calculated for each spot to derive a relative expression value. Spot ratio values were natural log transformed and normalized for interarray comparisons using the median-centered normalization method (12). Only spots that passed quality control criteria in ≥90% of all arrays were considered for statistical data analysis. Thus, the presented results were derived from a total of 5,552 filtered spots with reproducible and reliable gene expression measurements.

To identify genes with differential expression in the various experimental groups, we used the Statistical Analysis of Microarray Data (SAM) software (13). All statistical analyses were performed using both NSIs (NSI = CHI − CHB, i.e., background subtracted) and normalized median ratio channel signal (CHI). The data presented here are genes that were identified as statistically differentially expressed by both analyses. Normalized gene expression profiles were clustered using a Manhattan distance metric hierarchical clustering tool as implemented in the Institute for Genome Research Multiple Experiment Viewer (14). Class prediction by weighted vote method (15) was applied to find genes with the highest discriminative value in order to distinguish animals based on their phenotypic outcome (as was implemented in GeneSpring Software version 5.1).

Quantitative RT-PCR. Quantitative RT-PCR (QRT-PCR) analysis was performed using the Taqman 7900 HT Sequence Detection System (ABI Prism) apparatus with SYBR Green (Applied Biosystems) reagent and sequence-specific primer pairs (11). House keeping genes β-actin (ACTB) and hypoxanthine guanine ribotransferase (HPRT1) were used for normalization, since the expressions of these genes were relatively stable in our microarray experiment. The following primers were used: mouse CD36 5′ tgctggagctgttattggtg, 3′ catgagaatgcctccaaaca; mHSD3B4 5′ ggaaactgtgagcttcctgc, 3′ ggcggttaaatg gtggtcta; mKAP 5′ acagtctcctccggctttct, 3′ ctggattcccagctaatgga; mbeta actin (ACTB) 5′ accgtgaaaagatgatgacccag, 3′ agcctggatggctacgtaca; mHPRT 5′ tgttgttggatatgcccttg, 3′ ttgcgctcatcttaggcttt; mOPN 5′ tgatcaggacaacaacggaa, 3′ tctcctggctctctttggaa; human beta actin (ACTB) 5′ gatgagattggcatggcttt, 3′ caccttcaccgttccagttt; hHPRT1 5′ aaaggaccccacgaagtgtt, 3′ tcaagggcatatccta caacaa; and hOPN 5′ gccgaggtgatagtgtggtt, 3′ tactggatgtcaggtctgcg. After amplification, the melting curve analysis was performed to verify specificity of the amplifications.

Immunohistochemistry. Anti-mouse osteopontin antibody (MPIIIB10) was obtained from the Developmental Studies Hybridoma Bank, University of Iowa (Iowa City, IA) (16). Rabbit anti-mouse hydroxysteroid dehydrogenase-3β isotype 4 (HSD3B4) primary antibody was generated using a unique peptide of HSD3B4 by Sigma-Genosys and was affinity purified. Mouse anti-synaptopodin antibody was from Progen (Queensland, Australia), rabbit anti-synaptopodin antibody was a kind gift of Dr. Peter Mundel (Albert Einstein College of Medicine) (17), and rabbit anti-NaK/2Cl antibody was purchased from Chemicon (Temecula, CA). Immunostaining on paraffin-embedded tissue was performed as reported previously (18,19). Antibody specificity was verified via tissue incubation with the antigenic peptide/antibody mixture prepared 24 h before use.

Immunofluorescence labeling of frozen sections was performed with fluorescein and Cy3-labeled anti-mouse and anti-rabbit goat secondary antibodies (Jackson Immunochemical) as described previously (18). Images were captured using a digital charge-coupled device camera system connected to a Nikon microscope.

Human kidney biopsy samples. Sections of human kidney tissue were obtained from anonymous archived paraffin-embedded samples. All diabetic samples were from patients with biopsy-proven advanced diabetic nephropathy with a serum creatinine value in the range of 4–7 mg/dl and nephrotic range proteinuria. Institutional review board approval was obtained for procurement of kidney specimens at Thomas Jefferson University.

RESULTS

Classification of diabetic mice based on phenotype analysis. Our study included type 2 diabetic *db/db* mice (n = 46) and nondiabetic *db/m* controls (n = 33) on normal or high-protein animal chow. For generation of experimental type 1 diabetes, 8-week-old male C57BL/6J or 129 SvJ mice were given STZ or vehicle. All *db/db* mice and all STZ-induced diabetic mice developed comparable levels of hyperglycemia (Table 1). Albumin excretion rates in diabetic mice were significantly increased compared with those of controls and were similar in *db/db* and STZ-treated mice on normal chow (Table 1). Albuminuria was further increased in *db/db* mice on high-protein chow. Body and kidney weights are summarized in Table 1. At 24 weeks, glomeruli in *db/db* animals were characterized by diffuse mesangial matrix expansion, consistent with previous reports (20). In contrast, STZ-induced C57BL/6J diabetic mice developed mild (score 2) mesangial matrix expansion at 24 weeks (after approximately the same duration of diabetes as the *db/db* mice). Mesangial matrix expansion was more pronounced in 129SvJ than C57BL/6J mice but was not significantly increase by diabetes (Table 1). Our phenotype analysis identified hyperglycemia and mesangial matrix scores as phenotypic outcomes that allowed independent classification of animals as either hyperglycemic versus normoglycemic or advanced mesangial matrix expansion versus minimal/mild mesangial expansion. In this experiment all hyperglycemic mice had

TABLE 1
Phenotypic characterization of experimental animals

Genotype	*n*	Age (weeks)	Diet	Glucose (mg/dl)	Albuminuria (μg/24 h)	Mesangial matrix score	Body weight (g)	Kidney weight (mg)
db/m	8	16	Normal	151 ± 12	3.6 ± 1.7	1	22.7 ± 0.2	164 ± 4
db/m	15	24	Normal	161 ± 12	7.6 ± 2.5	1.5 ± 0.1	30 ± 0.5	216 ± 5
db/m	10	24	High protein	135 ± 17	10.4 ± 2	1.2 ± 0.1	29.9 ± 0.7	226 ± 12
db/db	6	16	Normal	557 ± 23*	78 ± 30†	2.1 ± 0.3*	45.7 ± 1*	219 ± 7*
db/db	10	24	Normal	543 ± 37*	68 ± 15*	3.8 ± 0.1*	48.1 ± 2*	228 ± 5
db/db	30	24	High protein	531 ± 13*	206 ± 36*	3.2 ± 0.1*	45 ± 0.8*	241 ± 6
C57BL/6J control	5	24	Normal	233 ± 4.5	25 ± 0.7	1	21.2 ± 0.7	138 ± 7
C57BL/6J STZ	8	24	Normal	516 ± 36*	69 ± 8*	2*	20.8 ± 0.5	197 ± 10†
129 SvJ control	5	30	Normal	111 ± 4	34 ± 12	2.5 ± 0.2	26 ± 0.6	221 ± 11
129 SvJ STZ	8	30	Normal	512 ± 27*	65 ± 7†	3.1 ± 0.3	23 ± 0.6	232 ± 8

Data are means ± SE. Bold face groups were included in microarray analysis. *$P < 0.01$ and †$P < 0.05$ compared with corresponding controls.

increased albumin excretion rates over baseline; these outcomes were therefore analyzed together.

Identification of genes differentially expressed in kidneys of STZ-induced diabetic C57BL/6J and *db/db* mice. To identify genes with statistically significant differential expression in the kidneys of 24-week-old *db/db* mice ($n = 29$) compared with control *db/m* mice ($n = 11$), the SAM software (13) was used. We identified 343 probes that satisfied our statistical threshold criteria (20% difference in mean ratio values and <1% false discovery rate [FDR]) as differentially expressed (Fig. 1*A*). Expression profiles of these genes were analyzed for coregulation and visualized by hierarchical clustering (Fig. 1*C*). Annotated lists of all the differentially expressed genes are available in the online data supplement Table 1 (available from http://diabetes.diabetesjournals.org). Similar analysis comparing 24-week-old STZ-induced diabetic C57BL/6J mice ($n = 5$) and vehicle-treated control mice ($n = 5$) identified 312 spots as differentially expressed (Fig. 1*A* and *B*) (online supplemental data Table 2). Of the 343 differentially expressed probes in *db/db* mice and 312 differentially expressed transcripts in the STZ-induced diabetic mouse kidneys, only 40 transcripts were differentially expressed in kidneys of both experimental diabetes models (12% of *db/db* mice and 13% of STZ-induced diabetic mice overlap). The small number of overlapping genes suggests that statistical comparisons based on a priori experimental group designs in two different well-established experimental models of diabetic kidney disease could identify a large number of genes that may not be related to diabetic kidney disease; their differential expression might rather be the result of other confounding variables, i.e., obesity, insulin level, genotype, etc.

Identification of differentially expressed genes based on statistical comparisons of groups with similar phenotype outcome. To identify differentially expressed genes that are associated with common phenotypic characteristics of diabetic kidney disease, experimental groups were separated into animals with hyperglycemia/albuminuria (29 *db/db* and 5 STZ-induced diabetic mice), as defined by blood glucose >300 mg/dl and albuminuria >30 μg/day, or with normoglycemia and absence of albuminuria (11 *db/m* and 5 control C57BL/6J mice). A total of 297 spots were thus identified as differentially expressed by SAM analysis (<1% FDR and 20% change) (Fig. 2*A*). Next, we classified mice into two groups according to the degree of mesangial matrix expansion, where animals with scores 3 or 4 (moderate/severe) (29 *db/db* mice) were compared with animals with scores 1 or 2 (minimal/mild) (5 STZ-induced diabetic, 5 vehicle-treated C57BL/6J, and 11 *db/m* mice), and identified 611 spots as differentially expressed by the SAM analysis (Fig. 2*A*). Annotated lists of these genes are available in the online supplementary data Table III*A* and *B*. Moreover, a large fraction of genes (63 and 39%) differentially expressed in hyperglycemic mice were also associated with the mesangial matrix expansion phenotype regardless of the genotype of the animals. The large fraction of commonly differentially expressed genes associated with the two phenotypic outcomes have a high likelihood of representing genes that are regulated by hyperglycemia/albuminuria and are involved in the development of mesangial disease.

To gain insight into the functional role of the differentially expressed genes, we annotated each gene target with their corresponding biological function, according to Gene Ontology Consortium data available in the Gencarta database system (Compugen, Tel Aviv, Israel). The data in each subgroup are represented as a proportion of total number of differentially expressed genes (Fig. 2*B*). This approach was used because the total numbers of differentially regulated transcripts were different in each comparison (297 in the hyperglycemia/albuminuria comparison and 611 in the mesangial matrix expansion comparison). Similarly, all of the analyzed genes from the microarray ($n = 5{,}552$) were annotated with their corresponding biological function. Genes associated with substrate and energy metabolism constituted the single largest functional group (~50% of all genes), and the representation in the differentially expressed groups was similar.

Next, statistical algorithms were applied to define significant differences in the differentially expressed groups compared with all of the analyzed genes (21). Genes annotated to cell proliferation and response to external stimuli functions were represented in statistically significantly different percentile among the differentially regulated genes (compared with all of the analyzed genes). This observation might underlie the importance of cell cycle dysregulation in diabetic kidney disease (22). The response to external stimuli ontology group includes number of cytokine-related genes, including several in the transforming growth factor-β (TGF-β) pathway. The

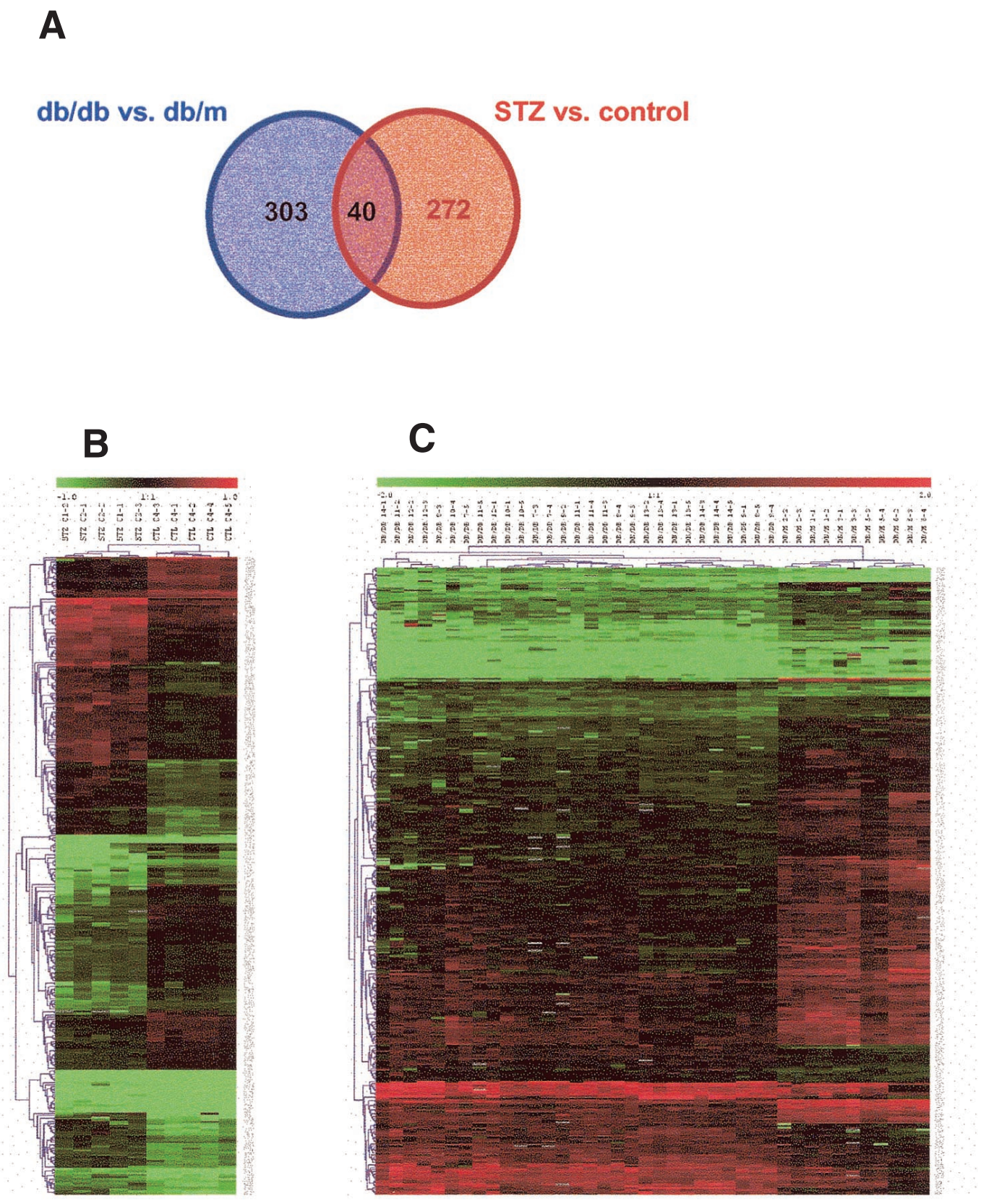

FIG. 1. Differentially expressed genes in kidneys of STZ-induced diabetic C57BL/6J or of *db/db* mouse models. *A*: Venn's diagram showing the overlap of differentially expressed genes by SAM analysis (1% FDR and 20% change in expression) comparing *db/db* and *db/m* mice (blue circle) and STZ-induced diabetic and control mice (red circle). Cluster analysis of differentially expressed genes comparing STZ-treated and control/untreated mice (*B*) and *db/db* and *db/m* mice (*C*). All data were mean centered, natural log transformed, and clustered using Pearson's correlation and Euclidian distance and average linkage. Each raw represents the expression profile of one transcript, and each column represents the profile of one animal. Red squares: transcripts that are overrepresented compared with the reference sample. Green squares: transcripts that are underrepresented compared with the reference sample. Gray: missing data values.

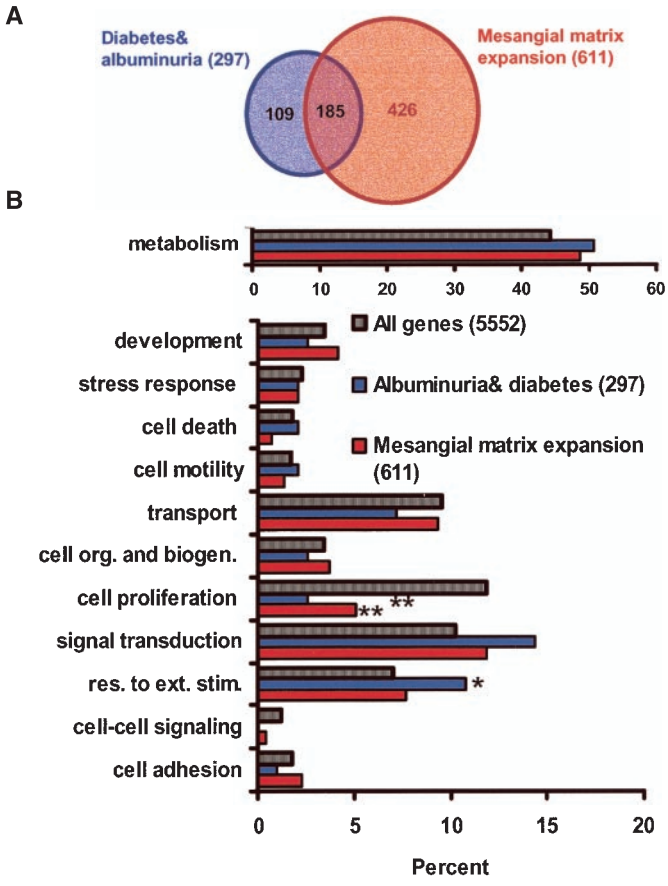

FIG. 2. Differentially expressed genes based on phenotypic manifestation. *A*: Genes that are differentially regulated in animals with diabetes and albuminuria (blue circle) and in animals with mesangial matrix expansion (red circle) and their overlap. *B*: Differentially expressed genes identified in each phenotypic outcome group were annotated with their biological function by GenCarta (Compugen) database. The distribution of genes in each gene ontology group and the normal distribution is shown. *$P < 0.05$ and **$P < 0.01$.

expression profile and the list of the genes involved in cell proliferation and response to external stimuli are shown in the online supplemental data, Fig. 1, in a cluster format.

Identification of candidate classifier genes for phenotypic outcomes in diabetic mouse kidneys. To decrease the number of candidate genes with expression profile characteristic to hyperglycemic/albuminuric mice or to mice with advanced glomerulosclerosis versus control, we used a supervised statistical class predictor algorithm (15). We selected a minimal set of genes that can provide optimal classification accuracy together with the highest prediction strength.

Initially 25 genes were selected whose expression values in the kidneys can discriminate hyperglycemic and albuminuric animals from control mice. The hierarchical cluster analysis of these genes classifying 50 experimental kidney RNA samples is shown in Fig. 3*A*. Repetitions of this analysis using progressively fewer predictor genes resulted in the identification of three genes that provided optimal classification of the origin of kidney RNA samples from hyperglycemic/albuminuric mice. The expression profile of these genes correctly identified 49 of 50 cases in leave-one-out cross validations. These genes were CD36, kidney androgen–regulated protein (KAP), and an unnamed expressed gene (GB#AA276424). Relative expression values of CD36, KAP, and the expressed sequence tags were all lower in kidneys of hyperglycemic/albuminuric compared with control mice.

A similar approach was used to identify 25 genes with the highest predictive values for discrimination of kidney RNA samples from mice with advanced mesangial matrix expansion (Fig. 3*B*). Expression data from two genes, *HSD3β4* and osteopontin (*OPN*), were sufficient to predict the origin of kidney RNA samples from mice with or without mesangial matrix expansion correctly in all 50 animals in cross-validation experiments. Relative mRNA expression values of *HSD3β4* were lower and *OPN* levels were higher in kidneys of mice with advanced mesangial expansion (Fig. 3*B*). We found a significant overlap in the list of the 25 top genes to classify mesangial matrix expansion and hyperglycemia/albuminuria. This might be partially due to the overlapping phenotype (e.g., all mice with mesangial disease are diabetic); however, since only a small number of genes (two or three) were sufficient to classify phenotypic outcome, it is likely that the other identified genes are confounders. This is demonstrated by the observation that the classification accuracy of the top 10 genes was lower (84%) than the top two genes (100%) to discriminate mice with advanced mesangial matrix expansion.

Validation of CD36, KAP, HSD3β4, and OPN as classifier genes. To confirm the relative transcript levels of CD36, KAP, HSD3β4, and OPN derived from the microarray experiment, we designed QRT-PCR assays to determine their transcript levels as an independent method of expression profiling. Figure 4 shows the distribution of the relative expression values obtained by QRT-PCR for the four identified classifier genes in diabetic and nondiabetic mice. There was a reduction of CD36, KAP, and HSD3β4 expression, whereas OPN was increased in the diabetic mice, thus confirming the gene expression results obtained by microarray for these genes. Receiver operator curve (ROC) analysis of the QRT-PCR data was used to determine the cutoff values for relative transcript levels of each gene that discriminated phenotype groups with highest specificity and high sensitivity (these values are shown on Fig. 4*A–D*).

To determine whether the observed expression profiles are also valid as classifiers using a group of test animals with different age, sex, and genetic background, we determined the transcript levels of our putative predictor genes in total kidney RNA samples from an independent series of 30-week-old 129 SvJ mice given STZ or vehicle or in 16-week-old female *db/db* mice (see Table 1 for phenotype description) that were not included in our initial microarray analysis. By applying the ROC-derived cutoff values (Fig. 4), we evaluated the performance of the relative expression levels of *CD36* and *KAP* in correctly classifying mice with hyperglycemia/albuminuria (Fig. 5*A* and *B*) and of *HSD3β4* and *OPN* in identifying mice with advanced mesangial matrix expansion (Fig. 5*C* and *D*). The lower relative gene expression level of *HSD3β4* was able to diagnose the presence or absence of mesangial matrix expansion in 72% of the animals (Fig. 5*C*). The relative expression of OPN performed better and was able to make the correct diagnosis (the presence or absence of ad-

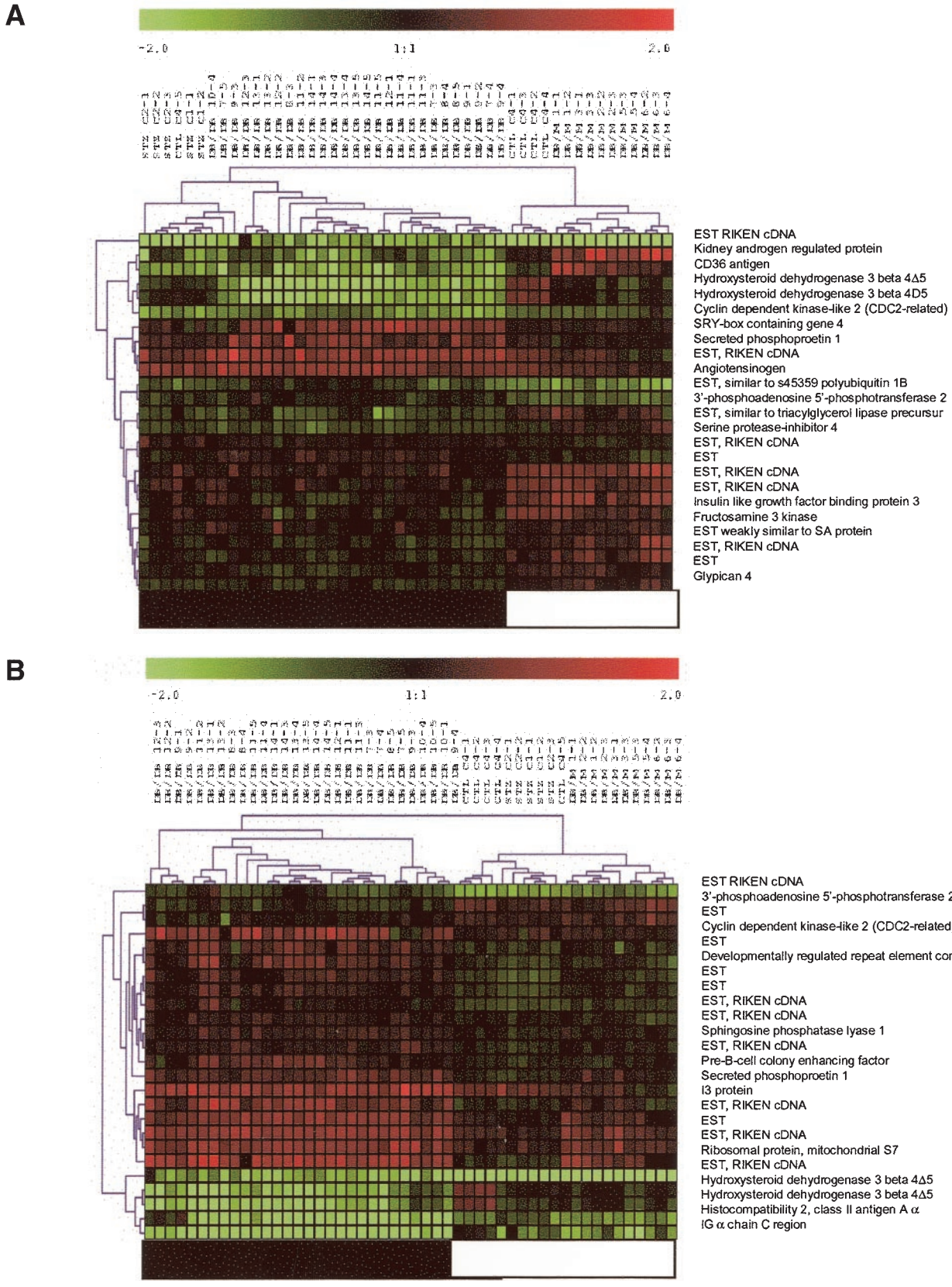

FIG. 3. Genes with the highest discriminative value to classify phenotypic outcome in mice. Hierarchical cluster analysis of the top 25 genes that can classify animals based on diabetes (*A*) and mesangial matrix expansion (*B*). All data were mean centered, natural log transformed, and clustered using Manhattan distance and complete linkage. Red squares: overrepresented transcripts compared with the reference sample. Green squares: underrepresented transcripts (compared with the reference RNA). Gray, missing data values. Animals with mesangial matrix expansion (*B*) are shown with a black bar and control animals with a white bar. Animals with diabetes (*A*) are represented with a black bar and control animals with a white bar.

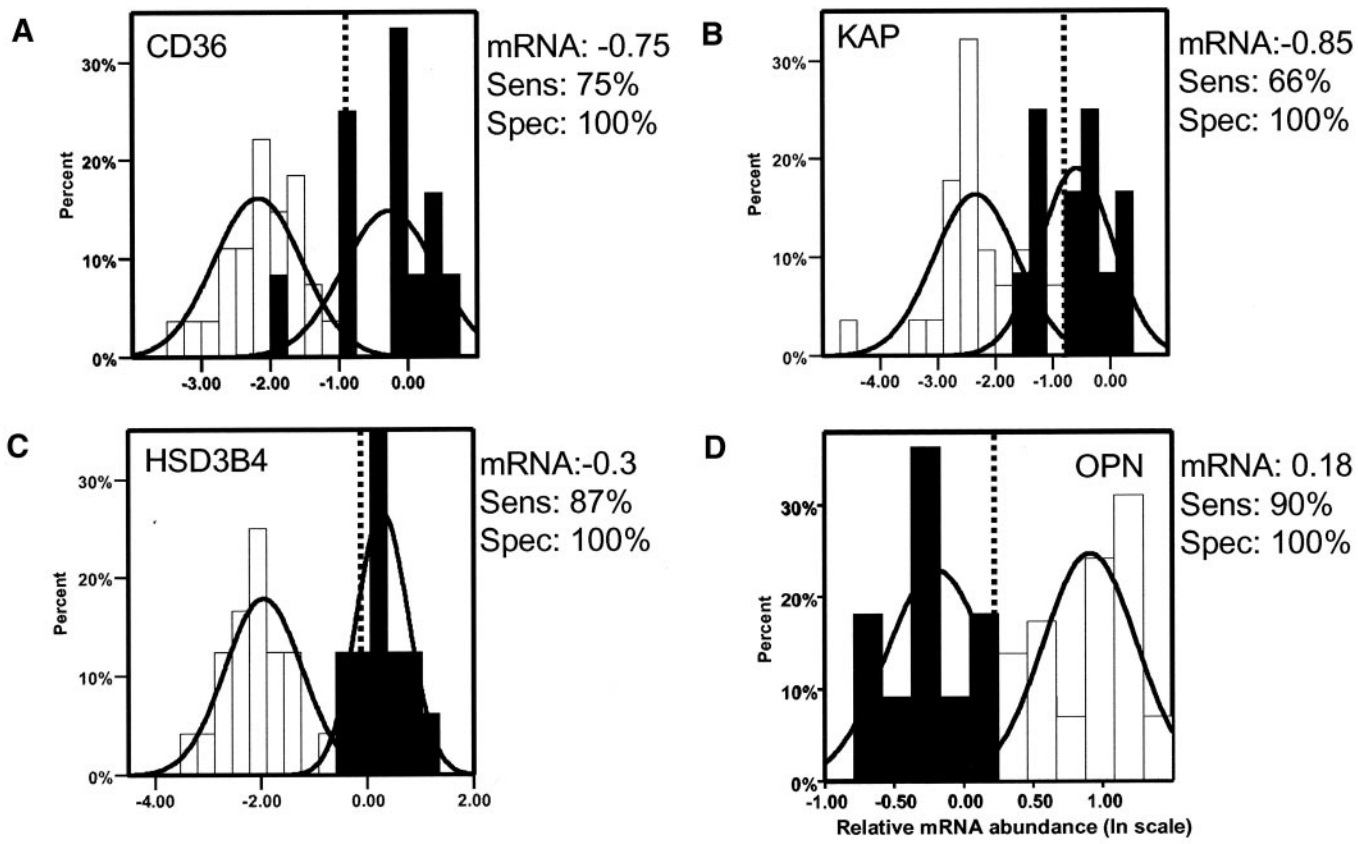

FIG. 4. Validation of the gene expression profile of the phenotype classifies genes by QRT-PCR. The histogram of the distribution of the relative expression of predictor genes in 50 mouse kidneys included in microarray analysis. The distribution of relative gene expression in control animals is shown in black. The CD36 (*A*) and KAP (*B*) distributions in animals with hyperglycemia are shown in white. The distribution of the expression of HSD3B4 (*C*) and OPN (*D*) in animals with mesangial matrix expansion is also represented in white. The black curves show the fitted normal distribution curves to the data. All relative expression data were compared with the same control (and normalized to housekeeping genes); all data are natural log transformed. The dotted lines correspond to the relative mRNA abundance values derived from ROC analysis. The relative values and corresponding sensitivities (Sens) and specificities (Spec) are shown on the right.

vanced mesangial matrix expansion) in 19 of 22 cases (Fig. 5*D*). In the classification of animals based on albuminuria and diabetes, CD36 performed better than KAP, and the prediction was correct in 95.5% of cases (Fig. 5*A* and *B*). Taken together, our results validate the data obtained from the microarray experiment and confirmed that mRNA levels of CD36 and KAP are strong classifiers of hyperglycemia, while HSD3β4 and OPN are the best predictors for mesangial expansion in different type and strains of diabetic mice, respectively.

Immunohistochemical localization of genes with high predictive value for mesangial matrix deposition. We performed immunohistochemical analysis of HSD3β4 and OPN to determine the protein expression and regulation patterns in normal and diabetic mouse and human kidneys. HSD3β4 is an enzyme that has been linked to testosterone metabolism, and isoform 4 was primarily expressed in the kidney (23); however, its cellular distribution in the kidney has not been characterized. We found specific HSD3β4 staining in podocytes and tubular cells in both control and diabetic human kidney samples (Fig. 6*A* and *B*). We observed similar staining pattern in kidney sections from *db/db* and *db/m* mice (data not shown). We confirmed the podocyte-specific expression pattern by double immunostaining with synaptopodin, a podocyte specific marker (Fig. 6*C–E*). The glomerular staining of HSD3β4 was decreased in human and murine diabetic samples with advanced glomerulosclerosis (Fig. 6*B*) compared with control (Fig. 6*A*). In summary, HSD3β4 was expressed in the tubules and in the glomerular podocytes in human and murine diabetic kidney samples, and its decreased expression correlated with the lower mRNA expression observed in the microarray experiment.

In control murine and human samples, we observed strong osteopontin expression in the medullary tubules (Fig. 7*A* and *B*). OPN in the murine samples colocalized with NaK/2Cl cotransporter staining, which is a specific marker of the loop of Henle (Fig. 7*F*) (24). We did not observe a change in the tubular expression of OPN between animals with diabetic glomerulopathy and controls (Fig. 7*A* and *B*). In addition to the tubular staining there was strong upregulation of OPN in the glomeruli of 24-week-old *db/db* mouse kidneys (Fig. 7*C* and *D*). We confirmed the podocyte-specific expression pattern again by double immunostaining with synaptopodin, a podocyte-specific marker (Fig. 7*E*). Similar to the murine kidney sections, we also observed OPN staining in podocytes in human diabetic kidney samples (Fig. 7*G*), which was completely blocked with the blocking peptide (Fig. 7*H*). In summary, immunohistochemical analysis showed that OPN expression in the renal tubules was not altered, while its expression in podocytes was strongly upregulated in diabetic kidney samples.

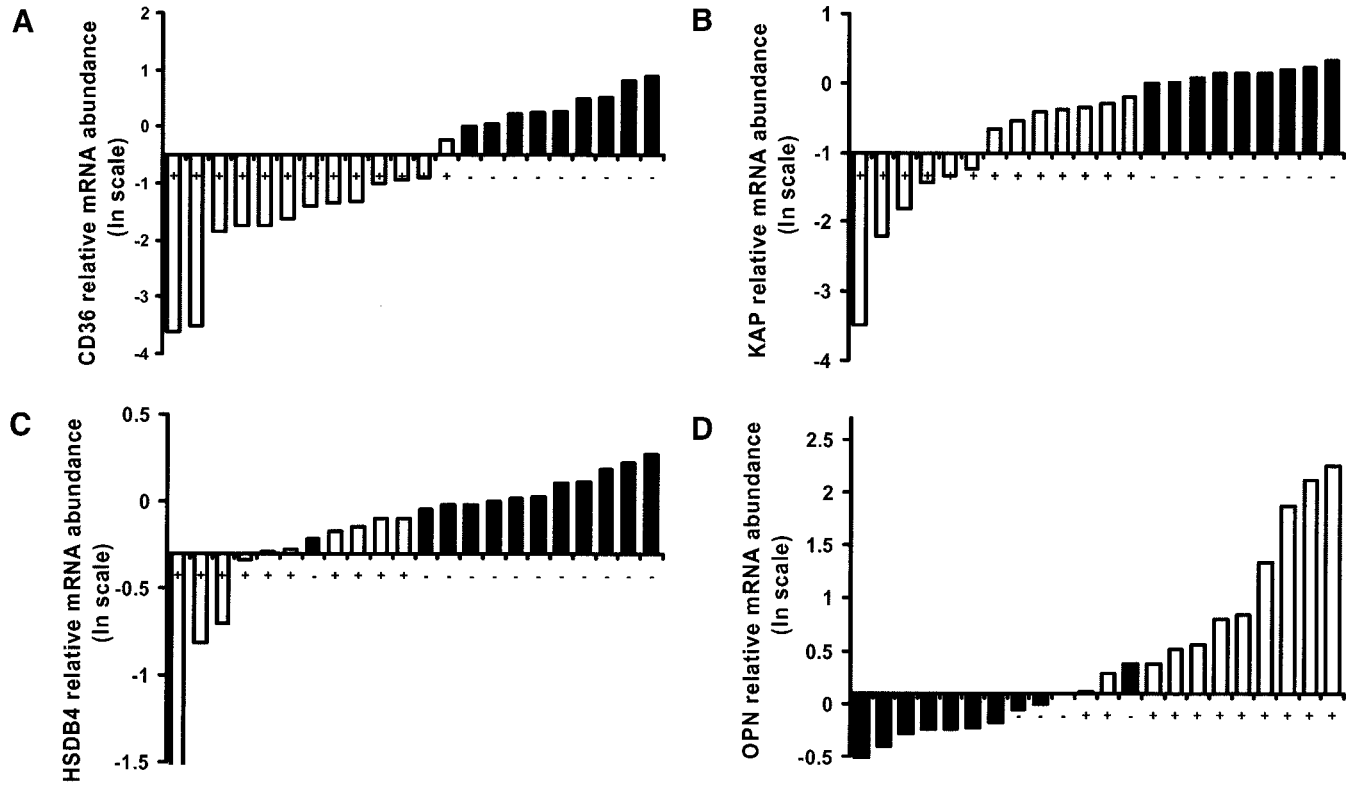

FIG. 5. Phenotype can be predicted in new samples based on gene expression profile. The relative expression of CD36 (*A*), KAP (*B*), HSD3B4 (*C*), and OPN (*D*) is shown in 22 mice not included in the microarray analysis. Each bar represents the relative gene expression in one animal. *A* and *B*: white bars (+) represent animals with diabetes and black bars represent control animals (-). *C* and *D*: white bars (+) represent animals with mesangial matrix expansion and black bars represents control mice (-). The *x*-axis crosses the *y*-axis at the relative expression value that was derived from the ROC analysis (see text and Fig. 4 for details).

DISCUSSION

Here we report gene expression profile analysis in various models and strains of murine diabetes. The inclusion of different models and strains allowed for the identification of genes whose expression is associated with specific steps of the diabetic complication regardless of the type of diabetes and the presence or absence of obesity, STZ, or leptin levels. We used two complementary approaches to identify genes with expression changes that are characteristic—and therefore more likely to be involved in the pathogenesis—of diabetic renal disease. First grouping animals for statistical comparisons based on phenotypic outcome provided a list of 611 differentially expressed probes in animals with diabetic glomerulopathy when compared with those with normal histology. In addition, a novel supervised neighborhood analysis was applied to

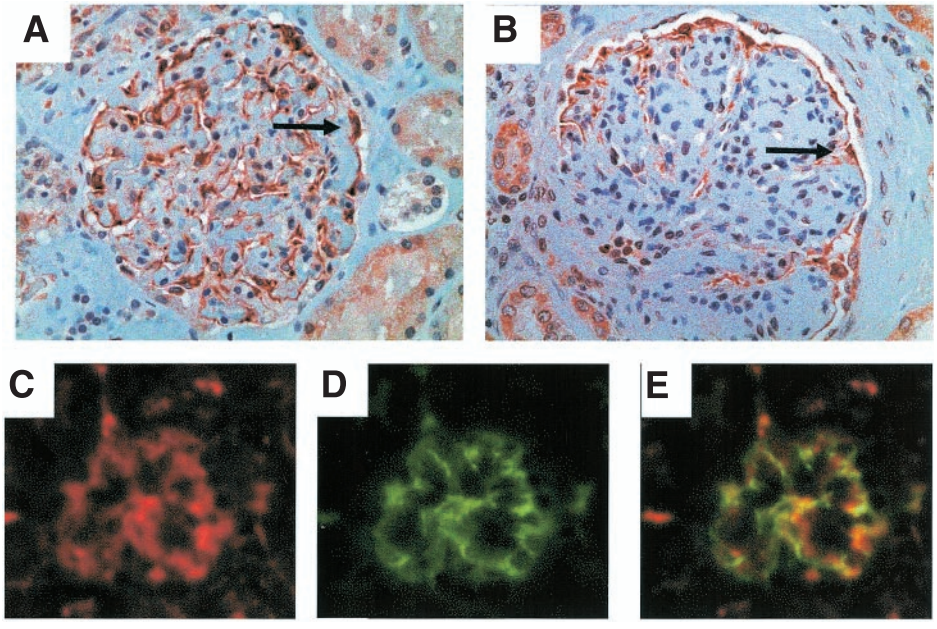

FIG. 6. Immunohistochemical localization of HSD3B4 in human and mouse control and diabetic kidneys. Podocytes and tubular epithelial cells stain positive (brown) with rabbit anti-HSD3B4 antibody in control human kidney tissue (*A*) and in kidneys with advanced diabetic nephropathy (*B*). Nuclei were counterstained with hematoxylin. HSD3B4 colocalized with podocyte marker synaptopodin in mice (*C–E*). Frozen control (4-week-old 129 SvJ) mouse kidney sections were stained with anti-HSD3B4 (red) (*C*) and anti-synaptopodin (green) (*D*) and their overlap (*E*). The arrows show positive HSD3B4 staining in podocytes.

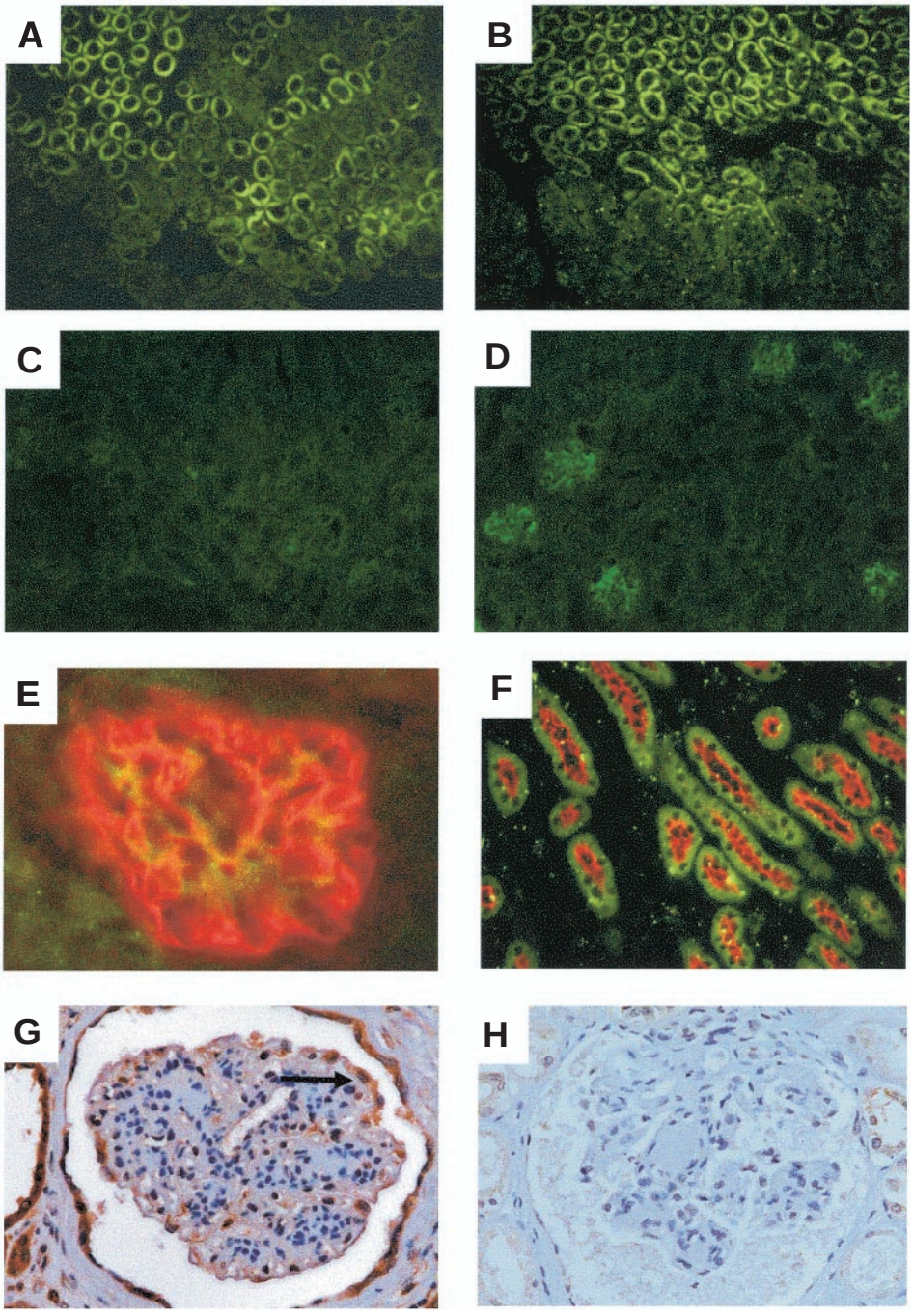

FIG. 7. Immunohistochemical localization of OPN in mouse control and diabetic kidneys. Normal (24-week-old *db/m*) (*A*, *C*, and *F*) and diabetic (24-week-old *db/db*) (*B*, *D*, and *E*) mouse kidney sections were stained with anti-OPN antibody and visualized with fluorescein-labeled goat anti-mouse secondary antibody (green). Sections *E* and *F* were costained with rabbit anti-synaptopodin and rabbit anti-Na/K/2Cl, respectively, and were developed with Cy3 goat anti-rabbit secondary antibody (red). The expression was increased in podocytes in diabetic samples (*D*) compared with the control tissue (*C*). Similar to murine samples, podocyte express OPN in human diabetic kidney samples (*G*), which can be blocked with neutralizing peptide (*H*). The arrow shows positive DPN staining in podocytes.

identify genes with patterns of expression that can classify animals with different phenotypic outcome. Since mesangial matrix expansion is a critical determinant for the development of diabetic nephropathy in humans, we applied this method to find genes with high discriminative value for diabetic glomerulopathy.

We identified HSD3β4 and OPN as lead genes in the mesangial matrix expansion signature. We were also able to categorize unknown samples into phenotypic groups based on relative gene expression values that we obtained via ROC analysis. This was particularly important since in the original microarray cohort, only *db/db* mice developed significant mesangial disease. The relevance of our findings is further supported by the observation that the distribution and regulation of OPN and HSD3β4 in human diabetic kidney disease followed a pattern similar to that we observed in the mouse models.

To our knowledge this is the first description of HSD3β4 distribution in the kidney with a specific antibody. The HSD family of enzymes characteristically catalyzes the conversion of Δ5-3-β hydroxysteroids by dehydrogenation of the 3-β-hydroxyl group and isomerization of the C5-C6 double bond to a C4-C5 double bond (23). However, HSD3β4 is unique in this family because it can catalyze the conversion of dihydrotestosterone to 5α-androstanediol in the presence of the cofactor NADPH, thus inactivating dihydrotestosterone (23). The localization of HSD3β4 to podocytes suggests that podocytes might be involved in testosterone inactivation and may thus be relevant to recent studies in which the administration of dihydrotes-

tosterone to rats with allografts led to marked proteinuria and glomerulosclerosis (25). As proteinuria and glomerulosclerosis are closely linked to podocyte dysfunction, it is likely that dihydrotestosterone results in deleterious effects on podocytes. In the context of diabetes, a decreased level of HSD3B4 may lead to enhanced dihydrotestosterone levels in the glomerulus and contribute to podocyte dysfunction, proteinuria, and glomerulosclerosis. These hypotheses will be examined in future studies.

We observed increased OPN expression in podocytes of diabetic mice with mesangial matrix expansion. We also found that OPN is expressed in vitro in cultured podocytes, which might underlie our in vivo observation (data not shown). Moreover, hyperglycemia and TGF-β increase OPN transcript levels in cultured podocytes (data not shown). It is conceivable that OPN might have a role in the cytoskeletal rearrangements of podocytes, since in cultured podocytes OPN coating induced a motile phenotype (26). OPN-mediated cytoskeletal rearrangement might be important in the development of diabetic nephropathy, where podocytes are under significant stress due to glomerular enlargement and/or podocyte loss (27), in which case the remaining podocytes must cover a larger area of the glomerulus. In contrast to previous studies showing increased tubular expression of OPN in STZ-induced diabetic rats (28), we did not find significant change in the tubular OPN expression in diabetic mice. This difference might be species specific, since OPN has also been shown to have a somewhat different tubular distribution in rats (29). Studies using OPN gene–deficient mice and antisense-treated animals have demonstrated that OPN promotes accumulation of macrophages and may play a role in macrophage-mediated renal injury (30). The observation might be important in explaining the presence of diabetic tubulointerstitial disease in diabetic rats and the complete absence of macrophage infiltration in the examined mouse models of diabetic nephropathy (E.B., K.Su., and K.Sh., unpublished observation) (30). On the other hand, there are reports suggesting renoprotective actions for OPN in various acute renal injury models (31).

Of interest is the observation that both of the proteins (HSD3β4 and OPN) that are closely linked to mesangial matrix expansion are present only in glomerular podocytes and not in glomerular endothelial or mesangial cells. This suggests an important cell-cell communication between podocytes and mesangial cells that results in the phenotype of mesangial matrix expansion.

In conclusion, the phenotype-based gene expression profile analysis provides a valuable framework to link variations in gene expression to phenotypic outcome. We have used this approach to identify an unbiased set of genes whose expression levels can identify mice with the development of diabetic glomerulopathy. The future determination of the predictive value of HSD3β4 and OPN for the development of diabetic nephropathy in humans might provide a new diagnostic tool and likely lead to a better understanding of the complex pathomechanism of diabetic nephropathy.

ACKNOWLEDGMENTS

K.Su. was supported by the Kidney and Urology Foundation and by the AECOM Clinical Investigator Fellowship. Studies were supported in part with funds from National Institutes of Health Grants R01 DK53867 (to K.Sh.) and U01 DK60995-01 (to E.B. and K.Sh.).

REFERENCES

1. *Renal Data System*: *USRDS 1998 Annual Data Report*. Washington, DC, U.S. Govt. Printing Office, 1998 (NIH publ. no. 98-3176)
2. Seaquist ER, Goetz FC, Rich S, Barbosa J: Familial clustering of diabetic kidney disease: evidence for genetic susceptibility to diabetic nephropathy. *N Engl J Med* 320:1161–1165, 1989
3. Susztak K, Sharma K, Schiffer M, McCue P, Ciccone E, Bottinger EP: Genomic strategies for diabetic nephropathy. *J Am Soc Nephrol* 14:S271–S278, 2003
4. Guttmacher AE, Collins FS: Genomic medicine: a primer. *N Engl J Med* 347:1512–1520, 2002
5. Perkins BA, Ficociello LH, Silva KH, Finkelstein DM, Warram JH, Krolewski AS: Regression of microalbuminuria in type 1 diabetes. *N Engl J Med* 348:2285–2293, 2003
6. Caramori ML, Fioretto P, Mauer M: The need for early predictors of diabetic nephropathy risk: is albumin excretion rate sufficient? *Diabetes* 49:1399–1408, 2000
7. Wada J, Zhang H, Tsuchiyama Y, Hiragushi K, Hida K, Shikata K, Kanwar YS, Makino H: Gene expression profile in streptozotocin-induced diabetic mice kidneys undergoing glomerulosclerosis. *Kidney Int* 59:1363–1373, 2001
8. Wilson KH, Eckenrode SE, Li QZ, Ruan QG, Yang P, Shi JD, Davoodi-Semiromi A, McIndoe RA, Croker BP, She JX: Microarray analysis of gene expression in the kidneys of new- and post-onset diabetic NOD mice. *Diabetes* 52:2151–2159, 2003
9. Clarkson MR, Murphy M, Gupta S, Lambe T, Mackenzie HS, Godson C, Martin F, Brady HR: High glucose-altered gene expression in mesangial cells: actin-regulatory protein gene expression is triggered by oxidative stress and cytoskeletal disassembly. *J Biol Chem* 277:9707–9712, 2002
10. Murphy M, Godson C, Cannon S, Kato S, Mackenzie HS, Martin F, Brady HR: Suppression subtractive hybridization identifies high glucose levels as a stimulus for expression of connective tissue growth factor and other genes in human mesangial cells. *J Biol Chem* 274:5830–5834, 1999
11. Bottinger EP, Novetsky A, Zavadil J: RNA labeling and hybridization of DNA microarrays. *Methods Mol Med* 86:275–284, 2003
12. Colantuoni C, Henry G, Zeger S, Pevsner J: Local mean normalization of microarray element signal intensities across an array surface: quality control and correction of spatially systematic artifacts. *Biotechniques* 32:1316–1320, 2002
13. Tusher VG, Tibshirani R, Chu G: Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* 98:5116–5121, 2001
14. Quackenbush J: Computational analysis of microarray data. *Nat Rev Genet* 2:418–427, 2001
15. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES: Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286:531–537, 1999
16. Gorski JP, Griffin D, Dudley G, Stanford C, Thomas R, Huang C, Lai E, Karr B, Solursh M: Bone acidic glycoprotein-75 is a major synthetic product of osteoblastic cells and localized as 75- and/or 50-kDa forms in mineralized phases of bone and growth plate and in serum. *J Biol Chem* 265:14956–14963, 1990
17. Weins A, Schwarz K, Faul C, Barisoni L, Linke WA, Mundel P: Differentiation- and stress-dependent nuclear cytoplasmic redistribution of myopodin, a novel actin-bundling protein. *J Cell Biol* 155:393–404, 2001
18. Schiffer M, Bitzer M, Roberts IS, Kopp JB, ten Dijke P, Mundel P, Bottinger EP: Apoptosis in podocytes induced by TGF-beta and Smad7. *J Clin Invest* 108:807–816, 2001
19. Schiffer M, von Gersdorff G, Bitzer M, Susztak K, Böttinger EP: Smad proteins and transforming growth factor-beta signaling. *Kidney Int Suppl* 77:S45–S52, 2000
20. Sharma K, McCue P, Dunn SR: Diabetic kidney disease in the db/db mouse. *Am J Physiol Renal Physiol* 284:F1138–F1144, 2003
21. Doniger S, Salomonis N, Dahlquist K, Vranizan K, CLawlor S, Conklin B: MAPPFinder: using Gene Ontology and GenMAPP to create a global gene-expression profile from microarray data. *Genome Biology* 4:R7.1–R7.12, 2003
22. Wolf G: Cell cycle regulation in diabetic nephropathy. *Kidney Int Suppl* 77:S59–S66, 2000
23. Payne AH, Abbaszade IG, Clarke TR, Bain PA, Park CH: The multiple

murine 3 beta-hydroxysteroid dehydrogenase isoforms: structure, function, and tissue- and developmentally specific expression. *Steroids* 62:169–175, 1997
24. Xie Y, Sakatsume M, Nishi S, Narita I, Arakawa M, Gejyo F: Expression, roles, receptors, and regulation of osteopontin in the kidney. *Kidney Int* 60:1645–1657, 2001
25. Antus B, Yao Y, Liu S, Song E, Lutz J, Heemann U: Contribution of androgens to chronic allograft nephropathy is mediated by dihydrotestosterone. *Kidney Int* 60:1955–1963, 2001
26. Endlich N, Sunohara M, Nietfeld W, Wolski EW, Schiwek D, Kranzlin B, Gretz N, Kriz W, Eickhoff H, Endlich K: Analysis of differential gene expression in stretched podocytes: osteopontin enhances adaptation of podocytes to mechanical stress. *FASEB J* 16:1850–1852, 2002
27. Steffes MW, Schmidt D, McCrery R, Basgen JM: Glomerular cell number in normal subjects and in type 1 diabetic patients. *Kidney Int* 59:2104–2113, 2001
28. Fischer JW, Tschope C, Reinecke A, Giachelli CM, Unger T: Upregulation of osteopontin expression in renal cortex of streptozotocin-induced diabetic rats is mediated by bradykinin. *Diabetes* 47:1512–1518, 1998
29. Khan SR, Johnson JM, Peck AB, Cornelius JG, Glenton PA: Expression of osteopontin in rat kidneys: induction during ethylene glycol induced calcium oxalate nephrolithiasis. *J Urol* 168:1173–1181, 2002
30. Kelly DJ, Wilkinson-Berka JL, Ricardo SD, Cox AJ, Gilbert RE: Progression of tubulointerstitial injury by osteopontin-induced macrophage recruitment in advanced diabetic nephropathy of transgenic (mRen-2)27 rats. *Nephrol Dial Transplant* 17:985–991, 2002
31. Rittling SR, Denhardt DT: Osteopontin function in pathology: lessons from osteopontin-deficient mice. *Exp Nephrol* 7:103–113, 1999