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 Sharma3

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 classifiers 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

A novel microarray strategy for the discovery of genes linked to diabetic kidney disease was introduced. This strategy includes a microarray analysis of murine and human kidney samples followed by immunohistochemical analysis of podocytes. The results of this study provide a new approach for the identification of gene expression changes that characterize the development of diabetic nephropathy.

© 2004 by the American Diabetes Association.
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**

Review of 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.

**RESULTS**

Classification of diabetic mice based on phenotype analysis. Our study included type 2 diabetic db/db mice (n = 40) 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. This experiment all hyperglycemic mice had...
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. 1A). Expression profiles of these genes were analyzed for coregulation and visualized by hierarchical clustering (Fig. 1C). 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. 1A 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 49 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. 2A). 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. 2A). Annotated lists of these genes are available in the online supplementary data Table IIIA 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 GeneCarta database system (Compugen, Tel Aviv, Israel). The data in each subgroup are represented as a proportion of total number of differentially expressed genes (Fig. 2B). 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 cellular 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.
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/albuninuric 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 from mice with advanced mesangial matrix expansion (Fig. 3B). 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. 3B). 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. 3B). 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. 4A–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 ourputative 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. 5A and B) and of HSD3β4 and OPN in identifying mice with advanced mesangial matrix expansion (Fig. 5C 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. 5C). 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.
vanced mesangial matrix expansion) in 19 of 22 cases (Fig. 5D). 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. 5A 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 HSD3B4 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 HSD3B4 and OPN to determine the protein expression and regulation patterns in normal and diabetic mouse and human kidneys. HSD3B4 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 HSD3B4 staining in podocytes and tubular cells in both control and diabetic human kidney samples (Fig. 6A 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. 6C–E). The glomerular staining of HSD3B4 was decreased in human and murine diabetic samples with advanced glomerulosclerosis (Fig. 6B) compared with control (Fig. 6A). In summary, HSD3B4 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. 7A and B). OPN in the murine samples colocalized with NaK/2Cl cotransporter staining, which is a specific marker of the loop of Henle (Fig. 7F) (24). We did not observe a change in the tubular expression of OPN between animals with diabetic glomerulopathy and controls (Fig. 7A 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. 7C and D). We confirmed the podocyte-specific expression pattern again by double immunostaining with synaptopodin, a podocyte-specific marker (Fig. 7E). Similar to the murine kidney sections, we also observed OPN staining in podocytes in human diabetic kidney samples (Fig. 7G), which was completely blocked with the blocking peptide (Fig. 7H). 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.

![Diagram of gene expression profile](image-url)
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. 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 represent 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).]

[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.]
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 dihydrotestosterone...
Podocytes are under significant stress due to glucerular enlargement and/or podocyte loss. This change might be important in the development of diabetic nephropathy, where podocytes are under significant stress due to glomerular enlargement and/or podocyte loss, 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, 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. 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. 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). On the other hand, there are reports suggesting renoprotective actions for OPN in various acute renal injury models.

Of interest is the observation that both of the proteins (HSD3B4 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 HSD3B4 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


23. Payne AH, Abbasszade IG, Clarke TR, Bain PA, Park CH: The multiple


