Kidney cytosine methylation changes improve renal function decline estimation in patients with diabetic kidney disease

Caroline Gluck, Chengxiang Qiu, Sang Youb Han, Matthew Palmer, Jihwan Park, Yi-An Ko, Yuting Guan, Xin Sheng, Robert L. Hanson, Jing Huang, Yong Chen, Ae Seo Deok Park, Maria Concepcion Izquierdo, Ioannis Mantzaris, Amit Verma, James Pullman, Hongzhe Li & Katalin Susztak

Epigenetic changes might provide the biological explanation for the long-lasting impact of metabolic alterations of diabetic kidney disease development. Here we examined cytosine methylation of human kidney tubules using Illumina Infinium 450 K arrays from 91 subjects with and without diabetes and varying degrees of kidney disease using a cross-sectional design. We identify cytosine methylation changes associated with kidney structural damage and build a model for kidney function decline. We find that the methylation levels of 65 probes are associated with the degree of kidney fibrosis at genome wide significance. In total 471 probes improve the model for kidney function decline. Methylation probes associated with kidney damage and functional decline enrich on kidney regulatory regions and associate with gene expression changes, including epidermal growth factor (EGF). Altogether, our work shows that kidney methylation differences can be detected in patients with diabetic kidney disease and improve kidney function decline models indicating that they are potentially functionally important.
Diabetes mellitus (DM) affects upwards of half a billion people worldwide and can lead to kidney complications, diabetic kidney disease (DKD), in ~40% of patients with DM\(^1\)\(^2\). The presence of kidney disease can explain most excess mortality associated with diabetes and DKD is the leading cause of end stage kidney disease worldwide\(^3\).

Despite the high clinical need and intense efforts, our understanding of DKD remains limited. Several factors explain this critical information gap. Animal models for diabetes do not develop renal complications, which limits their utility in studying the pathogenesis of DKD and in testing new drugs to treat DKD\(^4\)\(^5\). In addition, large-scale genetic efforts identified less than a handful of loci that reached genome wide significance for diabetic kidney disease, despite the strong heritability of DKD\(^6\)\(^7\). Finally, while there is increased risk of kidney disease development in patients with poor glycemic control, recent large interventional studies failed to show survival or renal benefit following normalization of blood glucose levels in patients with established diabetes\(^8\)\(^9\). These observations suggest that mechanisms outside of “traditional” genetic variants might be responsible for disease development.

Several groups have recently proposed that metabolic or developmental programming might play an important role in DKD development\(^10\)\(^-\)\(^15\). In the DCCT trial, patients initially assigned to the conventional glucose control group had an increased rate of DKD even after 25 years of strict glucose control, a phenomenon called metabolic memory\(^16\)\(^-\)\(^19\). Similarly, studies show consistent association between intrauterine nutritional deprivation and development of salt sensitive hypertension and kidney disease\(^20\)\(^-\)\(^27\). Epigenetic alterations could provide an explanation for these clinical observations. The epigenome is under the environmental influence and epigenetic changes are maintained during cell division. Therefore, metabolically driven epigenetic changes in the kidney could potentially explain the clinical observational data such as metabolic memory and developmental programming.

Several studies have tried to address the presence and role of epigenetic changes in DKD subjects. Changes in both histone and methylation patterns have been described when blood samples from patients with DM are compared to those with DKD\(^22\)\(^-\)\(^25\).\(^26\)\(^-\)\(^28\). While these early reports are interesting, most of these studies used small cohorts and most of the published changes failed to pass the genome wide statistical significance threshold. Furthermore, as epigenetic changes are cell type specific, it is not clear whether changes observed in blood samples correlate with kidney specific epigenetic differences. Our group has previously analyzed genome wide cytosine methylation changes in genomic DNA samples obtained from microdissected human kidney tubule samples of patients with mixed cohort of CKD\(^29\), hence the existence and role of epigenetic changes in DKD remains a critically important yet unanswered question.

While kidney function predictably declines with age, defining the speed of decline and detection of patients who will require renal replacement therapy is a critically important issue\(^30\)\(^-\)\(^32\). Patients with fast functional decline must be prioritized for intervention. Recent studies show that that baseline glomerular filtration rate (GFR), albuminuria and blood biochemical parameters can fairly accurately predict kidney function decline\(^33\). Additional biomarkers in the blood or urine have been identified recently that also predict renal function decline, however none has been shown to outperform the baseline clinical parameters. Epigenetic changes are the footprint of prior environmental alterations, but they are stable and inherited during cell division, therefore they could be ideal disease biomarkers.

The goal of the current project is to define genome wide cytosine methylation differences as measured by the Illumina Infinium 450k array in microdissected human kidney tubule epithelial cells of patients with diabetes and kidney disease. We integrate methylation changes with cell-type specific regulatory maps and gene expression changes to understand whether or not the observed changes are functionally important. Furthermore, we apply machine-learning algorithms to clinical and pathological descriptors to create renal function decline models and find that cytosine methylation levels can improve current models of renal function decline.

**Results**

**Methylation probes associate with degree of interstitial fibrosis.**

To describe epigenetic changes in diabetic kidney disease we analyzed microdissected human kidney tubule samples. As cytosine methylation is cell type specific, it is essential to study disease relevant cell types\(^34\). Our primary cohort included patients with and without diabetes and hypertension and varying degrees of diabetic kidney disease (DKD) as well as controls for comparison. The study used a cross sectional design and primary cohort sample size was 91. Demographics, clinical and histopathological analysis is shown on Tables 1 and 2. Histopathological analyses and quantitative scoring of tubulointerstitial fibrosis were performed by a renal pathologist. Principal component analysis for our primary cohort methylation data is available in Supplementary Figure 1.

Diabetic kidney disease has multiple manifestations, such as structure changes, mesangial expansion, glomerulosclerosis and tubulointerstitial fibrosis, in addition to functional changes of albuminuria and kidney function (eGFR) changes. The clinical definition of DKD includes albuminuria and GFR decline, however recent studies indicate a relatively poor correlation between these two manifestations. Structural changes are considered to be the gold standard to define DKD, therefore we used renal histological changes (mesangial expansion, fibrosis) to define DKD. The clinical (eGFR) and structural (fibrosis) descriptors showed a good, but not perfect correlation (Supplementary Table 1) and therefore they were analyzed independently. Furthermore, disease presentations follow a continuous pattern and this pattern was reflected in the analysis. Rather than using DKD as an outcome, we set to identify methylation changes associated with structural manifestation of DKD (i.e., tubulointerstitial fibrosis) or kidney function level (eGFR) using a linear regression model adjusted for key variables including age, sex, race, diabetes, hypertension, batch effect, bisulfite conversion efficiency, and degree of lymphocytic infiltrate on histology. To identify methylation changes that are not a consequence of genetic variation we have applied the Gap Hunter method and filtered out a large number of probes that were in the vicinity of regions with nucleotide variation. We found that the methylation level of 203 CpG probes significantly correlated with the degree of interstitial fibrosis using a linear regression with False Discovery Rate (FDR) adjusted \(p\)-value (FDR < 0.05) to determine genome-wide significance (Fig. 1a). Probes that showed significant methylation differences were distributed evenly across the genome (Fig. 1a). A similar number of probes showed lower methylation levels associated with interstitial fibrosis as showed higher methylation levels (Fig. 1b). The top probe that associated with kidney function level (eGFR) had a \(p\)-value of 1.02e−6 using linear regression (Supplementary Fig. 2).

**Validation of fibrosis-associated methylation probes.** We sought to validate the methylation changes associated with
Again, we used a linear regression analysis while adjusting for age, and 3. Principal component analysis for our replication kidney validation dataset can be found under Supplementary Tables 2, Demographics, clinical and histopathological descriptors of the chronic kidney disease (CKD), as well as healthy controls. 85 samples from patients with mixed diabetic and hypertensive kidney disease therefore we used an independent cohort of NATURE COMMUNICATIONS | https://doi.org/10.1038/s41467-019-10378-8 ARTICLE Vessel arteriolar hyalinosis (0–3) 0.07 (0.26)–Vessel intimal thickening (0–3) 0.68 (0.75)–Vessel medial thickening (0–3) 0.35 (0.78)–Mesangial matrix (0–3) 0.35 (0.78)–Glomerular wall thickening (0–3) 0.25 (0.66)–Mesangial matrix (0–3) 0.35 (0.78)–Hypoperfused glomeruli (0–3) 0.07 (0.26)–Segmentally sclerotic glomeruli (%) 10.41 (15.66)–Globally sclerotic glomeruli (%) 10.41 (15.66)–Tubular atrophy (%) 12.52 (21.03)–Acute tubular injury (%) 1.27 (5.05)–Tubules reabsorption (O–3) 0.21 (0.46)–Interstitial fibrosis (%) 12.21 (18.59)–Plasmacytic infiltrate (O–3) 0.31 (0.56)–Lymphocytic infiltrate (O–3) 0.78 (0.75)–Eosinophilic infiltrate (O–3) 0.16 (0.43)–Vessel medial thickening (O–3) 0.07 (0.26)–Vessel intimal fibrosis (O–3) 1.23 (0.89)–Vessel arteriolar hyalinosis (O–3) 0.43 (0.78)

Data are mean (SD) or n (%) fibrosis, as an important manifestation of DKD. There is a significant overlap in fibrotic changes in diabetic and hypertensive kidney disease therefore we used an independent cohort of 85 samples from patients with mixed diabetic and hypertensive chronic kidney disease (CKD), as well as healthy controls. Demographics, clinical and histopathological descriptors of the validation dataset can be found under Supplementary Tables 2 and 3. Principal component analysis for our replication kidney cohort methylation data is available in Supplementary Figure 3. Again, we used a linear regression analysis while adjusting for age, sex, race, diabetes, hypertension, batch effect, bisulfite conversion efficiency, and degree of lymphocytic infiltrate on histology. Again, probes potentially influenced by genotypic variation were filtered out using the fairly conservative Gap Hunter method. Of the 203 CpG probes identified in the primary dataset, we replicated the association (p-value < 0.05) for 65 CpG probes using linear regression with directional consistency of the methylation change. In the combined cohort, methylation level of all 65 CpG probes were associated with interstitial fibrosis using linear regression and stringent significance criteria for multiple comparisons, FDR corrected p-value < 0.05, corresponding to nominal p-value < 3.23E–05 (Fig. 2). In summary, methylation level at 65 CpG probes were associated with interstitial fibrosis (independent of CKD etiology) and replicated in an independent cohort using strict statistical correction for multiple comparisons (Supplementary Data 1). Gene ontology and Ingenuity Pathway Analysis (IPA) for replicated methylation probes associated with degree of interstitial fibrosis is shown in Supplementary Table 4 and Supplementary Figure 4, respectively, and included genes associated
with cell adhesion, localization, cell death and survival, and connective tissue development and function.

**Fibrosis-associated probes on gene regulatory regions.** Next, we wanted to understand whether methylation changes observed in kidney tubule cells could be functionally important. This follow-up analysis was performed using the 65 probes for which methylation changes were replicated in an independent cohort. Methylation changes that are localized to regions where transcription factors can bind are more likely to be functionally important as they can alter transcriptional accessibility and downstream gene expression levels. We took advantage of the fact that regulatory regions are enriched for specific histone tail modifications to define transcriptionally active regions in the adult human kidney. Healthy human adult kidney ChIP-Seq data was obtained through the Roadmap Epigenomic Project. Using the ChromHMM algorithm we have integrated signals for H3K4me1, H3K4me3, H3K27ac, H3K36me3, H3K9ac, and H3K9me3 data to identify promoter, enhancer and other regulatory regions in the human kidney. We mapped the location of the 65 CpG probes for which methylation level was associated with interstitial fibrosis. As compared with all probes of the HumanMethylation450k array used in the primary analysis \((n = 321,473)\), the 65 CpG probes associated with interstitial fibrosis were enriched in kidney enhancer regions \( \beta > 0.5075 \). The analysis identified 1791 significant methylation-expression correlation pairs (Supplementary Data 2). Since we were interested in methylation changes that are functional in kidney cells, we narrowed our list by looking only at CpG probes located in active kidney regulatory (promoter, enhancer, transcribed) regions. Probes that were significantly associated with interstitial fibrosis and confirmed independent replication \((n = 65)\) were mapped to identify the narrowest set of probes with the highest probability to play a functional role. Five probes were identified that passed these criteria and their association with fibrosis, gene expression and genomic location is shown in Table 3. For top probe, cg20597486, we completed experiments to validate the methylation changes as measured by the Illumina Infinium 450 K arrays (Supplementary Figure 6). Correlations between all top probes associated with degree of interstitial fibrosis and nearby (Cis) gene expression changes is shown in Supplementary Figure 7. Figure 3 further describes one of the top probes: cg20597486. Cg20597486 methylation level was significantly associated with interstitial fibrosis. This probe was located in a kidney promoter region, as shown in Fig. 3b locus zoom indicating that it is likely to be functionally important. While the probe correlated with gene expression changes of multiple nearby genes, it is closest to IFI16 transcription start site (TSS), therefore, this gene is most likely to be influenced by methylation changes at this location. For probes that are further away from the gene TSS, functional link would require further experimental confirmation. Probe methylation level correlated with IFI16 (Gamma-interferon-inducible protein) transcript level in microdissected human kidney tubule samples. These results highlight that decrease of the methylation at cg20597486 is likely to be important for kidney disease, as it is associated with an increase in IFI16 expression.

**Probes in regulatory regions associate with gene expression.** In order to further support the functional importance of the methylation of the 65 CpG probes in kidney disease development, we correlated methylation changes with gene expression levels analyzed in the same kidney tubule samples. Not all samples had available Affymetrix gene expression data and the demographics and clinical description of this sub-cohort is available in Supplementary Table 5. We correlated CpG probe methylation level change with nearby \((500 \text{ kb})\) gene expression changes. We used random permutation method to determine significance between methylation and gene expression and used a cutoff \( p \text{-value} = 8 \times 10^{-5} \). Using this \( p \)-value, we had > 80% power to detect an effect size of \( \beta > 0.5075 \). The analysis identified 1791 significant methylation-expression correlation pairs (Supplementary Data 2). Since we were interested in methylation changes that are functional in kidney cells, we narrowed our list by looking only at CpG probes located in active kidney regulatory (promoter, enhancer, transcribed) regions. Probes that were significantly associated with interstitial fibrosis and confirmed independent replication \((n = 65)\) were mapped to identify the narrowest set of probes with the highest probability to play a functional role. Five probes were identified that passed these criteria and their association with fibrosis, gene expression and genomic location is shown in Table 3. For top probe, cg20597486, we completed experiments to validate the methylation changes as measured by the Illumina Infinium 450 K arrays (Supplementary Figure 6). Correlations between all top probes associated with degree of interstitial fibrosis and nearby (Cis) gene expression changes is shown in Supplementary Figure 7. Figure 3 further describes one of the top probes: cg20597486. Cg20597486 methylation level was significantly associated with interstitial fibrosis. This probe was located in a kidney promoter region, as shown in Fig. 3b locus zoom indicating that it is likely to be functionally important. While the probe correlated with gene expression changes of multiple nearby genes, it is closest to IFI16 transcription start site (TSS), therefore, this gene is most likely to be influenced by methylation changes at this location. For probes that are further away from the gene TSS, functional link would require further experimental confirmation. Probe methylation level correlated with IFI16 (Gamma-interferon-inducible protein) transcript level in microdissected human kidney tubule samples. These results highlight that decrease of the methylation at cg20597486 is likely to be important for kidney disease, as it is associated with an increase in IFI16 expression.

**Building a model for kidney function decline.** In order to further elucidate methylation changes with increased likelihood of being causal for kidney fibrosis development, we examined probes whose methylation were associated with the rate of future kidney function decline. Multiple factors have already been identified that can predict kidney function decline. Our goal was to understand whether or not cytosine methylation changes independently associate with functional decline (i.e., they are less likely to be a proxy for a known clinical variable). In order to do this, we first defined the association between clinical variables and CKD progression. This analysis was performed in the primary cohort subset with longitudinal eGFR measurements \((n = 69)\) (Clinical information is available Supplementary Table 5). CKD
progression was defined by eGFR slope, which was adjusted to account for random variation as well as non-normality (Supplementary Figure 8). Low baseline eGFR and higher interstitial fibrosis was associated with faster eGFR decline (more negative adjusted eGFR slope) as has been described previously (Supplementary Table 6, Supplementary Figure 9). We started with all available clinical and histological variables that correlated with adjusted eGFR slope on univariate analysis (p-value < 0.05) using a two-sided Pearson correlation test (Supplementary Table 7). We then selected variables using a machine learning regression analysis method, least absolute shrinkage and selection operator (LASSO), in order to improve model accuracy and reduce model overfitting. In our dataset, the following parameters showed significant (independent) association with the rate kidney function decline: baseline eGFR, diabetes status, and age. These parameters overlap with the published variables known to predict progression was de...
kidney disease, such as eGFR, age and albuminuria validating our dataset (Supplementary Table 8).

While hypertension has been shown to predict CKD progression in some cohorts, we examined if the addition of hypertension to our CKD model would alter our outcome. In our data, only the diagnosis of HTN significantly correlated with adjusted eGFR slope on univariate analysis using a two-sided Pearson correlation test ($beta = -0.29$ $p$-value $= 0.017$); systolic blood pressure and mean arterial pressure (MAP) were not ($beta = -0.07$, $p$-value $= 0.22$; and $beta = 0.13$, $p$-value $= 0.32$, respectively) (Supplementary Table 7). Our data is consistent with results of large collaborative cohorts that created kidney function decline prediction models, such as Tangri et al. These predictive models did not include hypertension or blood pressure diagnosis for functional decline prediction yet were able to reach a C-statistics score of around 0.939.

### Gene expression does not improve CKD progression model

Once we confirmed the clinical parameters that predicted functional decline we added genome wide gene expression changes to the linear regression model. All linear regression models were run as a weighted regression whereby subjects with increased variability to their rate of decline were weighted less. Gene expression levels failed to significantly improve our precision to predict kidney function decline at genome wide significance (FDR $< 0.05$) using a weighted linear regression analysis. Our analysis indicates that failure to improve prediction was not due to a reduced power as the gene expression dataset was smaller ($n = 58$) than the methylation dataset.

### Cytosine methylation improves CKD progression model

Next, we tested whether epigenetic changes can improve kidney function decline prediction. We individually added CpG probe methylation level to the base CKD progression model. The methylation level of 471 probes was significantly associated with renal function decline (adjusted eGFR slope) at genome wide significance (FDR $< 0.05$) and improved model fit as measured by model AIC (AIC $< 206$) (Supplementary Data 3) using a weighted linear regression analysis. Gene ontology annotation for methylation probes that improved our CKD progression model is shown in Supplementary Table 9 and Ingenuity Pathway Analysis in Supplementary Figure 10. We found only minimal overlap between methylation probes that improved kidney function decline and those that correlated with fibrosis (only 2 probes of 65) (Supplementary Table 8). If we force hypertension into our model, we find that 67 of the 471 methylation probes still significantly improved the model (FDR $< 0.05$) (Supplementary Data 4) using a weighted linear regression analysis. Our results show that kidney tubule methylation changes do improve renal function decline models beyond the already established parameters.

### Validation of top progression-associated methylation probes

Methylation changes that predict kidney function decline and improve upon existing prediction models are of critical importance therefore we set to validate these results in an independent cohort. Unfortunately, we did not have additional methylation data obtained from human kidney tubule samples with longitudinal eGFR data. On the other hand, we had access to an independent genome wide methylation dataset of peripheral blood mononuclear cells from a well phenotyped cohort of American Indians with detailed longitudinal follow-up that included measurement of eGFR in examinations subsequent to the sample in which methylation was measured. This cohort was also analyzed by Illumina 450k arrays and we ensured that data processing was performed using a similar pipeline (see Methods). Principal component analysis for this cohort is available in Supplementary Figure 12. Cytosine methylation changes associated with adjusted eGFR slope were determined by a weighted linear regression analysis and controlled for variables such as age, sex, duration of diabetes, mean blood pressure, hemoglobin A1c, batch effect, bisulphite conversion efficiency as well as distribution of cell types. The methylation levels of 25 CpG probes that improved our CKD progression model fit in the kidney also were significantly correlated with kidney function decline in this independent cohort that analyzed blood samples ($p$-value $< 0.05$ with consistent direction of methylation change) using a weighted linear regression analysis (Fig. 5).

Finally, we compared our results with those recently published by Chen et al. that reported changes in the methylome of blood samples of 63 patients with type1 diabetes form the DCCT/EDIC cohort. We identified some consistencies in the results as shown in Supplementary Data 6, even though we analyzed different cell types (blood vs. kidney), different types of diabetes (DCCT type1 vs. ours type2), and different analytical methods (use of Gap Hunter) and different outcomes (progressive retinopathy/nephropathy vs degree of interstitial fibrosis and eGFR decline). In summary, methylation changes associated with kidney fibrosis can be replicated in other kidney cohort samples and to lesser degree in blood samples.

Overall, DNA cytosine methylation levels of the renal tubules can improve kidney function decline estimations and we have identified specific methylation changes that are associated with kidney function decline.

### Discussion

To our knowledge this is the first study to describe cytosine methylation differences in kidney tubule samples of patients with DKD. Earlier reported studies used blood samples or analyzed
non-diabetic kidney disease.\textsuperscript{22,31–35,49} Most prior studies used modest sample sizes and most of these results failed to pass genome wide significance levels. Changes observed in renal tubule cells are robust compared to results published for blood samples for similar sample size.\textsuperscript{22,31–34,49} DKD has multiple manifestations and we examined functional and structural changes separately. It is interesting to note that the association was weaker with functional changes (eGFR) measured at a single time-point,
However, the association with structural and longitudinal changes were robust and reproducible. While functional and structural changes strongly correlate, our recent analysis has identified significant differences between functional and structural changes. We found that similar to methylation differences, gene expression changes correlate more strongly with structural changes than with kidney function. Given that eGFR is strongly determined by systemic blood pressure and volume status it does seem to make sense that gene expression and methylation changes have stronger correlation with kidney fibrosis. Renal tubule cells and fibrosis play key roles in DKD development and therefore it is plausible that the association observed in the cell type of interest is more proximal to disease development and stronger.

Our studies indicate not only the robustness of the association but also that methylation changes are likely to be functionally important as they are located on kidney specific regulatory regions and correlate with gene expression changes. Methylation changes are postulated to affect gene expression through interference of transcription factor binding at gene regulatory regions. This relationship is most evident at proximal regulatory regions such as promoters, however, methylation changes at distant regulatory regions (for example enhancers) may similarly affect transcription of nearby genes. In these instances, functional linkage would require experimental confirmation. Regardless, several identified methylation-gene expression pairs appear to be highly interesting. For example, in our samples, decreased methylation at cg20597486 is associated with increased percent interstitial fibrosis on kidney histology and increased expression of IFI16, Gamma-interferon-inducible protein. IFI16 is an important transcriptional regulator and may modulate NF-κB activation, which has been shown to play a role in kidney injury in mouse models. Gene ontology analysis shows enrichment of functional regulators could have significant impact on disease development or prenatal or adult programming. However, further testing will be needed to establish causality, for example in a mouse model.

Another key finding of the current work is the identification of methylation changes that can improve kidney function decline models. It is important to note that while the expression of more than a thousand genes correlated with kidney fibrosis at baseline in our study, gene expression levels failed to improve kidney function decline models. Cytosine methylation changes on the other hand were able to improve upon already existing models for kidney function decline, even though the clinical models show high and reproducible C-statistics.

We would also like to point out an interesting convergence of epigenetic and gene expression changes. Prior functional studies have identified a correlation between kidney EGF transcript and fibrosis and urinary EGF levels and kidney function decline. In our study we replicated the correlation of EGF and kidney fibrosis; however, only the methylation status of EGF, not the gene expression level, showed significant correlation with future kidney function decline even after adjustment for baseline GFR, age, and diabetes status. As cytosine methylation levels are relatively stable it makes them potentially ideal disease biomarkers. Gene expression changes follow a minute to minute regulation by environmental factors therefore they might show increased variability and less reproducibility as predictors of progression. The increased stability of DNA over RNA makes methylation levels a more attractive biomarker. Such a finding could have important clinical significance for the development of biomarkers and therapeutics for DKD and our understanding of the pathogenesis of DKD.

While we used microdissected human kidney tubule samples and controlled for as many known and hypothesized sources of variation (i.e., lymphocytic infiltrate) as we could, it is possible that additional unaccounted variation, such as changing cell proportions associated with kidney fibrosis, influenced our results. Despite all our efforts we cannot exclude the contribution of cell heterogeneity to the observed methylation changes. Future studies should aim to control for this by utilizing single-cell technology that currently exists for RNA sequencing but is not as mature for whole genome epigenome analysis. Additionally, while we aimed to narrow our results by focusing on probes located in active kidney gene regulatory regions, it is possible that these regions are different in the diseased state or in rare cell types such as fibroblasts. For example, we would like to highlight that a top probe cg21048700, that improved the CKD progression model and was associated with changes in COL3A1 (type III collagen) expression, was not located on regulatory region when whole kidney tubule samples analyzed, however, may still be integral to disease development as this probe is localized on a regulatory region in fibroblasts.

Another potential pitfall of our study is that our longitudinal eGFR samples are from subjects who underwent full/partial nephrectomies that could potentially affect their rate of kidney function decline. However, since all subjects underwent full/partial nephrectomy and we were analyzing relative rate of decline, this likely did not affect our results. In addition, most patients underwent nephrectomy for the indication of renal cell carcinoma, which typically is cured through resection without the

---

Table 5 Progression model with top probes associated with gene expression changes

| Variable | Base model | Base model | Base model | % Explained by variable | % Explained by variable | % Explained by variable |
|----------|------------|------------|------------|-------------------------|-------------------------|-------------------------|
| Baseline GFR | 0.03*** | 0.04*** | 0.03*** | 20.67 | 30.78 | 22.03 |
| Diabetes | –0.72 | –1.09*** | –1.17*** | 5.75 | 8.87 | 10.75 |
| Age | –0.03 | –0.01 | –0.01 | 0.16 | 0.16 | 0.11 |
| Cpg probe | NA | 1.24*** | –1.21*** | NA | 17.99 | 13.02 |
| Methylation batch | NA | NA | NA | 10.71 | 11.26 |
| Bisulfite conversion | NA | 20.28* | 17.70 | NA | 2.90 | 2.46 |
| K^2 | 0.51 | 0.76 | 0.70 | 0.49 | 0.70 | 0.63 |
| Adjusted K^2 | 0.49 | 0.70 | 0.63 | 206.1 | 177.1 | 1911 |
| Akaike information criterion | 3.13e–10 | 5.58e–13 | 1.23e–10 |

---

a For each variable, coefficient estimates are shown with the following significance codes: 0 ‘***’; 0.001 ‘**’; 0.01 ‘*’; 0.05 ‘.’

b Model is a weighted linear regression model of adjusted eGFR slope (weight = inverse variance of adjusted eGFR slope). Base model includes variables: baseline eGFR, Diabetes, and Age. Models D and E include base variables with the addition of methylation level at probe location, methylation batch, and bisulfite conversion efficiency.

c Proportion of variance explained by the variable based on conditional sums of squares calculated in Type II ANOVA analysis.

---

8 NATURE COMMUNICATIONS | https://doi.org/10.1038/s41467-019-10378-8
indicate that sequence variations could also drive cell type independent methylation changes which is in keeping with observations from other diseases36–39. Future studies shall aim to dissect the direct contribution of environmental factors and sequence variations in cytosine methylation in the kidney. Despite these limitations, here we identified pathways and replicated previously observed pathways associated with kidney disease pathogenesis and progression. Given the correlation with the gene expression pattern and the association with future functional decline it is likely that some of these methylation changes are also functionally important for DKD development.

Overall we show genome wide significant and replicated cytosine methylation differences in DKD and use machine learning methods to identify loci that can improve current models for kidney function decline. These loci should be studied in future prospective cohort studies as candidate biomarkers or pathways for intervention.

**Methods**

**Participants.** The primary cohort included 91 human kidney samples obtained from surgical full/partial nephrectomies. Kidney samples were ~0.5 cm in diameter and were surrounded by at least 2 cm of normal tissue margins. The study used cross-sectional design. The samples were collected from Albert Einstein College of Medicine Montefiore Medical Center between the years of 2007–2011. Samples were de-identified and the corresponding clinical information available at the time of nephrectomy was collected by an honest broker. Hypertension and diabetes diagnosis were based on chart review diagnosis codes. Blood pressure values were obtained at the time of tissue procurement. In this cohort, 41 subjects carried a diagnosis of DM and 65 carried a diagnosis of HTN and there were 22 patients with diabetic CKD. Estimated GFR was determined using the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) estimating equation. In this primary cohort, all subjects with GFR < 60 had diabetes and evidence of DKD on histologic analysis (n = 22). Subjects with GFR > 60 included those with and without diabetes and with and without hypertension. Clinical data were obtained at the time of the sample collection and for a subset of samples estimated glomerular filtration rate (eGFR) measures were available from clinical records before and after the nephrectomy (n = 69). Only subjects with longitudinal eGFR measurements for at least three months post nephrectomy were included for the analysis. The mean timespan of follow-up was 2.4 years (SD = 1.5 years).

Institutional Review Boards at the Albert Einstein College of Medicine Montefiore Medical Center (IRB 2002–2002) and the University of Pennsylvania (IRB 815796) reviewed this study. This project utilized de-identified kidney biospecimens. Therefore, this project does not meet the definition of human subject research and IRB review was not required. It was completed in compliance with all relevant ethical regulations.

The replication cohort included 85 microdissected human kidney tubule samples obtained from surgical full/partial nephrectomies. The cause of CKD in this cohort was mixed diabetic and hypertensive kidney disease. There were 10 patients with diabetic CKD and 38 patients with non-diabetic CKD (where CKD corresponds with eGFR < 60 ml/min per 1.73 m²). 11 patients had diabetes and 25 had hypertension in the absence of CKD. The replication data set included 23 samples from our primary data set in order to assess degree of batch effect. The second replication cohort was comprised of 115 peripheral blood samples obtained from American Indians with DKD enrolled in a longitudinal study described by Qiu et al. Blood samples for DNA isolation and cytosine methylation analysis were collected at the baseline examination in patients with diabetes and chronic kidney disease (albumin to creatinine ratio ≥ 300 mg/g or eGFR < 60 ml/min per 1.73 m²). The mean timespan of follow-up was 5.6 years (SD = 3.5 years). 45 cases reached the endpoint of end stage renal disease.

**Procedures.** A portion of the harvested kidney was formalin fixed and paraffin embedded and stored for Period Acid Schiff and Hematoxylin and Eosin staining. The histopathology was evaluated by a blinded pathologist and described using a descriptor method scoring 19 independent parameters including the degree of tubulointerstitial fibrosis. Histopathological lesions were used to determine the cause of CKD in our study. Seven samples in the primary and eight samples in the replication cohort had missing fibrosis score.

The rest of the kidney was freshly immersed into RNALater and stored at -80°C. Samples were manually microdissected into glomerular and tubular compartments under a stereomicroscope using established methods. DNA and RNA was then isolated from tubule tissue using the Qiagen RNAeasy and DNAeasy or AllPrep DNA/RNA mini kits following manufacturer’s instructions. Transcript level changes were determined using the Affymetrix U133 RNA microarray. DNA was bisulfite converted using the EZ DNA methylation kit (Zymo research) and assayed using the Illumina Infinium HumanMethylation450K BeadChip according to the manufacturer’s instructions. The bisulfit conversion efficiency was
calculated using the bisulfite conversion control probes, based on Illumina guidelines. Ten CpG sites designated by Illumina as control sites (6 CpGs targeted by the type 1 probes and 4 CpGs targeted by type II probes), where expected each CpG to be 100% methylated, were used to control for non-complete bisulfite conversion. The bisulfite conversion efficiency used in the primary analysis was the median methylation estimate from the ten control sites. The bisulfite conversion term was calculated by taking the median value of the probes that Illumina provided to estimate conversion efficiency97. Each cohort (primary and replication cohorts) was run as a single batch at different facilities. Therefore, they were initially analyzed separately and then combined to avoid capturing facility-batch batch artifact98. Of note, all analysis were additionally controlled for batch effect as defined by Illumina slide position (Sentrix ID).

DNA methylation values were extracted using the minfi package99. Our methylation analysis pipeline included several quality control measures. We used the Gap-Hunter package100 to remove CpG probes that are in the proximity of regions with genetic variations (gap probes). We also removed probes located on the sex chromosome and known to cross-hybridize to other loci. We also excluded probes with poor detection p value (p > 0.01) and control probes. Finally, we left with 321,473 CpG probes in the primary cohort. The same preprocessing was performed for the kidney replication cohort and resulted in 355,141 CpG probes in the kidney replication cohort. The blood replication cohort was processed in the same pipeline with the exception of application of the Gap-Hunter program instead of excluding with a cut-off. We normalized methylation data utilizing beta-mixture quantile normalization for the replication cohort). In order to account for probe bias inherent to the 450 K array, samples from the kidney replication cohort and two samples from the blood replication cohort. Principal component analysis was performed on each cohort (dbSNP137) were removed. Finally, there were 401,438 CpG probes in the blood replication cohort. Principal component analysis component was used to estimate the technical replicate effect and to establish the relationship between samples from different batches. LASSO selected variables were used to create a CKD risk score model. Weights for each variable were determined based on Akaike Information Criterion (AIC). CpG probe methylation was individually added to the base model using weighted linear regression (weight = inverse variance of adjusted eGFR slope) and additionally controlled for batch effect and bisulfite conversion efficiency. False discovery rate (FDR) procedure was used to account for multiple comparisons98. Methylxation Probes that improved model fit (AIC lower than base model) and reached epigenome wide statistical significance (FDR < 0.05) by weighted linear regression analysis were further analyzed. We estimated that with approximately 1:3 case to control ratio, that we had >80% power to detect mean differences of targeted genes and albuminuria in patients with early diabetic kidney disease in the AKita model. Weighted linear regression was also used to analyze the association between cytokine methylation and adjusted eGFR slope (weight = inverse variance of adjusted eGFR slope) in our blood replication cohort. Covariates in this regression model included age, sex, duration of diabetes, mean blood pressure, hemoglobin A1c, batch effect, bisulfite conversion efficiency as well as distribution of cell types100.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

**Data availability**

The methylation data is available at GEO: GSE50087. The expression data is available at ArrayExpress: E-MTAB-5929, E-MTAB-2502. The rest of the data are available from the corresponding author upon reasonable request.

**Received:** 22 November 2017 **Accepted:** 7 May 2019

**Published online:** 05 June 2019

---

**References**

1. Thomas, M. C. et al. Diabetic kidney disease. *Nat. Rev. Dis. Prim.* 1, 15018 (2015).
2. Thomas, M. C., Cooper, M. E. & Zimmet, P. Changing epidemiology of type 2 diabetes mellitus and associated chronic kidney disease. *Nat. Rev. Nephrol.* 12, 73–81 (2016).
3. Briosu, F. C. 3rd et al. Mouse models of diabetic nephropathy. *J. Am. Soc. Nephrol.* 20, 2503–2512 (2009).
4. Iyer, S. K. et al. Genome-wide association and trans-ethnic meta-analysis for advanced diabetic kidney disease: family investigation of nephropathy and diabetes (FIND). *PloS Genet.* 11, e1005352 (2015).
5. Palmer, N. D. et al. A genome-wide association search for type 2 diabetes genes in African Americans. *PloS ONE 7*, e92022 (2012).
6. Sandholm, N. et al. Chromosome 2q31.1 associates with ESRD in women with type 1 diabetes. *J. Am. Soc. Nephrol.* 24, 1537–1543 (2013).
7. Patarro, C. et al. Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. *Nat. Commun.* 7, 10023 (2016).
8. Consortium, T. G. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multivariate gene regulation in humans. *Science 348*, 648–660 (2015).
9. Sandholm, N. et al. New susceptibility loci associated with kidney disease in type 1 diabetes. *PloS Genet.* 8, e1002921 (2012).
10. Duckworth, W. et al. Glucose control and vascular complications in veterans with type 2 diabetes. *N. Engl. J. Med.* 360, 129–139 (2009).
11. Ismail-Beigi, F. et al. Effect of intensive treatment of hyperglycaemia on microvascular outcomes in type 2 diabetes: an ACCORD randomised trial. *Lancet 376*, 419–430 (2010).
12. Wyatt, C. M. & Cattran, D. C. Intensive glycemic control and the risk of end-stage renal disease: an ADVANCE in the management of diabetes? *Kidney Int.* 90, 8–10 (2016).
13. Marumo, T. et al. Diabetes induces aberrant DNA methylation in the proximal tubules of the kidney. *J. Am. Soc. Nephrol.* 26, 2388–2397 (2015).
14. Reddy, M. A. et al. Losartan reverses permissive epigenetic changes in renal glomeruli of diabetic db/db mice. *Kidney Int. 85*, 362–373 (2014).
15. Aldemir, O., Turgut, F. & Gokce, C. The association between methylation levels of targeted genes and albuminuria in patients with early diabetic kidney disease. *Ren. Fail* 39, 597–601 (2017).
16. Reddy, M. A., Zhang, E. & Natarajan, R. Epigenetic mechanisms in diabetic complications and metabolic memory. *Diabetologia 58*, 443–455 (2015).
17. Tikoo, K., Sharma, E., Amara, V. R., Pamulapati, H. & Dhawale, V. S. DNA methyltransferase 1 may be a therapy target for attenuating diabetic nephropathy and podocyte injury. *Kidney Int.* 92, 140–153 (2017).
18. Kaelin, W. G. Jr. & McKnight, S. L. Influence of metabolism on epigenetics and disease. *Cell 153*, 56–69 (2013).
19. Beckerman, P., Ko, Y. A. & Susztak, K. Epigenetics: a new way to look at kidney disease. *Nephrol. Dial. Transpl.* 29, 1821–1827 (2014).
20. Sayed, S. G. et al. Progressive glomerulosclerosis in type 2 diabetes is associated with renal histone H3K9 and H3K23 acetylation, H3K4 dimethylation and phosphorylation at serine 10. *Nephrol. Dial. Transpl.* 25, 1811–1817 (2010).
...
Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-019-10378-8.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Journal peer review information: Nature Communications thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.