5-Hydroxymethylcytosine Profiles in Plasma Cell-free DNA Reflect Molecular Characteristics of Diabetic Kidney Disease

Jin-Lin Chu

Xinjiang Medical University

Shu-Hong Bi

Peking University Third Hospital

Yao He

Peking University

Rui-Yao Ma

Xinjiang Medical University

Xing-Yu Wan

Peking University Third Hospital

Zi-Hao Wang

Beijing Institute of Pharmacology and Toxicology: Academy of Military Medical Sciences Institute of Pharmacology and Toxicology

Lei Zhang

Natural Medicine Institute of Zhejiang YangShengTang Co Ltd

Meng-Zhu Zheng

Peking University Third Hospital

Zhan-Qun Yang

Peking University Third Hospital

Ling-Wei Du

Hainan University

Yiminiguli Maimaiti

Xinjiang Medical University

Gulinazi Biekedawulaiti

Xinjiang Medical University

Maimaitiyasen Duolikun

Xinjiang Medical University

Hang-yu Chen

Peking University Third Hospital

Long Chen

Peking University Third Hospital

Lin-Lin Li (✉ lilinlin01@xjmu.edu.cn)

Xinjiang Medical University

Lu Tie

Peking University

Jian Lin

Peking University Third Hospital

Research Article

Keywords: diabetic kidney disease, epigenetics, 5-hydroxymethylcytosine, cell-free DNA

Posted Date: January 11th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1165504/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Complications of diabetes mellitus (DM) are the leading cause of DM-related disability and mortality. Notably, diabetic kidney disease (DKD), one of the main complications of DM, has become a frequent cause of end-stage renal disease. A clinically convenient, non-invasive approach for monitoring the development of DKD would benefit the overall life quality of patients with DM and contribute to lower medical burdens through promoting preventive interventions.

Methods: We utilized 5hmC-Seal to profile genome-wide 5-hydroxymethylcytosines in plasma cell-free DNA (cfDNA). Candidate genes were identified by intersecting the differentially modified 5hmC marker genes (DMGs) and differentially expressed genes (DEGs) from the GEO datasets GSE30528 and GSE30529. Cytoscape software was used to construct the protein-protein interaction (PPI) network and identify the hub genes.

Results: The final gene panel of 9 hub genes, including (CTNNB1, PTEN, MYD88, ITGAM, CD28, ITGB2, VCAM1, CXCR4, CD44) were confirmed. Further analysis indicated that this 9-gene signature showed a good capacity to distinguish between DKD and DM.

Conclusions: The 5hmC-Seal assay was successfully applied to the cfDNA samples from a cohort of DM patients with or without DKD. Altered 5hmC signatures in plasma cfDNA indicate that 5hmC-Seal has the potential to be a non-invasive epigenetic tool for monitoring the development of DKD and be a part of diabetic care.

Background

According to the latest report from the Diabetes Atlas of the International Diabetes Federation, the number of diabetic patients is expected to increase to 643 million by 2030 and 784 million by 2045. In 2021, diabetes has caused 6.7 million deaths, which means that one patient died every 5 seconds. Diabetic kidney disease (DKD) is one of the main microvascular complications of diabetes mellitus (DM) that remains a leading cause of end-stage renal disease (ESRD)[1]. Worth mentioning, 30-40% of patients with DM develop DKD[2]. Clinical examinations of proteinuria and estimated glomerular filtration rate are used to detect DKD[3]. However, many patients with type 1 diabetes, and most with type 2 diabetes, do not follow this classic course in modern clinical practice. For example, many diabetic patients with renal impairment do not manifest excessive urinary albumin loss[4, 5]. Discovering a set of non-invasive surrogate markers that reflect the progression of DM to DKD is urgently needed.

Previous publications have shown that genes related to the development of DKD are not only regulated by classical signaling pathways but also regulated via epigenetic mechanisms, including through chromatin histone modification, DNA methylation, and hydroxymethylation[6, 7]. 5-hydroxymethylcytosine (5hmC), the oxidative product of 5-methylcytosine (5mC) catalyzed by ten-eleven translocation enzymes, is a relatively stable intermediate of active DNA demethylation and is regarded as an important epigenetic feature[8]. Plasma cell-free DNA 5hmC profiling has been proved to be a powerful

tool in the early detection and progression monitoring of cancers[9, 10]. Plasma 5hmC profiling-based methods were also reported for the detection of vascular complications in DM patients [7, 11].

Herein, we set up a genome-wide, plasma cell-free DNA 5hmC profiling-based method for the distinguishment between DM and DKD patients. Plasma cfDNA from 17 DKD patients was subjected to 5hmC-Seal sequencing, and differential 5hmC markers associated with DKD were identified. The selected gene panel was further evaluated as biomarkers to identify DKD patients. This work may serve as preliminary research for the diagnosis of DKD using plasma cfDNA 5hmC marks.

Methods

Study participants

These samples of DM patients (n=14), DKD patients (n=17), and healthy controls (n=26) were recruited from the Peking University Third Hospital. Informed consent was obtained from all participants before the study. Diagnosis of DM is based on the Diabetes Medical Care Standards published in 2020 by the American Diabetes Association (ADA)[12]. Diagnostic criteria of DKD was according to the KDIGO 2020 Clinical Practice Guideline and the Consensus for Prevention and Treatment of Diabetic Kidney Disease 2014 of Chinese Diabetes Society and mainly includes urinary albumin-to-creatinine ratio (UACR) ≥ 30 mg/g and/or estimated glomerular filtration rate (eGFR) $<60 \text{ mL} \cdot \text{min}^{-1} \cdot (1.73\text{m}^2)^{-1}$, and lasts for more than three months[13]. Clinical data were collected from all samples, and detailed information is listed in Table 1.

Acquisition of microarray data

The mRNA expression and related experimental and clinical data of DKD were downloaded from GEO using the search terms diabetic kidney disease. The gene expression microarray datasets GSE30528 and GSE30529 were selected and downloaded. GSE30528 and GSE30529 were used for differentially expressed genes (DEGs) screening. Combined with the results of 5hmC-Seal, it was identified that the markers are also differentially expressed in DKD tissues. The detailed process of study design is shown in Fig 1.

Clinical samples collection and cfDNA isolation

Whole blood specimens were obtained by routine venous phlebotomy and collected into Cell-Free DNA Collection Tubes (Roche). Tubes were maintained at 15 °C to 25 °C with plasma separation performed within 24 h by centrifugation of whole blood at 1350×g for 15 min at 4 °C and 13,500×g for 5 min at 4 °C, followed by transfer of the plasma layer to a new tube. Plasma was aliquoted for subsequent cfDNA isolation or storage at -80 °C. The plasma cfDNA was extracted using the QIAamp Circulating Nucleic Acid Kit (QIAGEN) following the manufacturer's protocol and then stored at -20 °C.

5hmC library construction and high throughput sequencing

5hmC library construction for all samples was consistent with previously reported [14, 15]. The cfDNA samples were end-repaired using the KAPA Hyper Prep Kit (KAPA Biosystems) and then ligated to sequencing adapters. 5hmC bases were biotinylated via two-step chemistry and purified by the DNA Clean & Concentrator Kit (ZYMO), and subsequently enriched by binding to Streptavidin beads (Life Technologies). Then, the beads were resuspended in RNase-free water and amplified with PCR. Finally, the PCR products were purified using AMPure XP beads (Beckman). All libraries were quantified with a Qubit 3.0 fluorometer (Life Technologies). 5hmC sequencing was performed on the NextSeq 500 platform according to paired-end 39-bp high-throughput sequencing.

Mapping and identifying 5hmC enriched regions

FastQC (version 0.11.5) was used to check the sequence quality. Raw 5hmC-Seal data were aligned to the human genome reference (hg19) with bowtie 2 (version 2.2.9)[16]. Pair-end reads were extended and converted into BedGraph format and normalized the total number of aligned reads using Bedtools (version 2.19.1) [17], and then converted to bigwig format by using bedGraphToBigWig from the Integrated Genomics Viewer to visualize. Potential 5hmC enriched regions were identified using MACS2 (version 2.1.1) in each sample[18]. Peak regions that appeared in more than 10 samples and that were less than 1000 bp were combined into one unified catalog for each patient. Genomic regions that tend to show artifact signals, according to ENCODE, were filtered out. The 5hmC enriched regions were generated by intersecting the individual peak call file with the merged peak file. We used the CHIP seeker package to annotate the 5hmC-enriched regions, and genes that were closest to the regions were used for the following analysis.

Exploring functional relevance of DMGs

After filtering genes in chromosome X and Y, differentially modified genes (DMGs) in autosomes between DM and DKD samples were detected using the negative binomial generalized linear model in DESeq2 package ($|\log_2FC| > 0.5$ and $p\text{-value} < 0.05$). The enrichment analysis of the GO biological process (BP) was completed by the ClueGO (version 2.5.8) and CluePedia (version 1.5.8) plug-in from Cytoscape software (version 3.7.1). Medium network specificity, Bonferroni adjusted $p \leq 0.01$, and enriched gene number ≥ 5 were chosen as the criteria for significance. The pathway enrichment analyses of DMGs were performed using the Cluster Profile package in Bioconductor, and $p\text{-value} \leq 0.05$ was considered statistically significant[19].

Protein-Protein Interaction (PPI) construction and selection of hub genes

We used the STRING database (<https://string-db.org/>) to conduct PPI for DMGs[20]. We mapped the DMGs onto the PPI network and set an interaction score > 0.4 . Then, the Cytoscape software was used to visualize and construct the PPI network. To identify the key PPI network modules, the MCODE (version 1.6.1) plugin from the Cytoscape software was used to perform the gene network clustering analysis with the filter criteria of K-core=2, max depth=100, degree cut-off=2, and node score cut-off=0.2[21]. CytoHubba (version 0.1) was used to identify significant genes in the PPI network as hub genes[22]. We

used the MCC algorithm to calculate the top 10 genes. Finally, all the results were intersected to obtain the final hub genes.

Analysis of hub genes

Hub genes were analyzed by principal component analysis (PCA) using the FactoMineR package. Hierarchical clustering and heatmap analysis of DM and DKD on 31 samples were done by the pheatmap package. Correlation analysis between the clinical parameters and hub genes level was performed using Pearson correlation.

Statistical analysis

For clinical data, continuous data are presented as mean (SD), and categorical data are presented as numbers and percentages. Three groups were compared using one-way analysis of variance and Kruskal Wallis H test parameters and non-parametric continuous variables, while unpaired t-test and Mann Whitney U test parameters and non-parametric continuous variables were used for two groups comparisons. Subsequent analysis was carried out using R version 4.1.0.

Results

5hmC distribution in plasma cfDNA of DKD patients

Plasma samples from 14 DM patients, 17 DKD patients, and 13 healthy donors were collected (Table 1). 5hmC-Seal was performed with extracted cfDNA to map the genome-wide 5hmC profiles for all samples. Sequencing data showed that 5hmC was enriched within transcription start sites (TSS) and transcription termination sites (TTS) and depleted in the flanking regions (Fig. 2A), which was consistent with previous reports[10], suggesting that the accumulation of 5hmC is related to transcriptional activity. The total peaks detected per million reads in the DKD cfDNA cohort were significantly less than in DM and Control cohorts (Fig. 2B). In addition, the overall 5hmC enrichment was significantly different between the three groups (Control, DM, DKD) for any genomic feature types (Fig. 2C). Differential analysis between DM and DKD revealed 2069 hyper-hydroxymethylated genes and 3099 hypo-hydroxymethylated genes in DKD (Fig. 2D). The majority of these differential genes were enriched in intronic, intergenic, or promoter regions (Fig. 2E), which was consistent with previous reports[23]. The involving signaling pathways for the hyper-hydroxymethylated genes were classified in insulin secretion, carbohydrate digestion and absorption, and cysteine and methionine metabolism (Fig. 2F). While the pathways for hypo-hydroxymethylated genes were included in growth hormone synthesis, EGFR tyrosine kinase inhibitor resistance, and Axon guidance (Fig. 2G). A heatmap was generated by performing unsupervised clustering of the top 200 DMGs between DM and DKD (Fig. 2H).

Overlapping markers screening and function exploration

To examine the impact of DMGs in the glomerular and tubular tissue of DKD patients, the 5hmC profiles were compared among DKD tubular (n=10) and glomerular (n=9) samples. A total of 72 overlapping gene

markers were identified as kidney-related DMGs (Fig. 3A-B). One individual marker, CTNNB1, was randomly selected and tested with a set of randomly selected patients and healthy donors (Fig. 3C), which proved the reliability of the identified gene panels. Pathway analysis of 72 overlapping markers in DKD patients suggested functional enrichment in certain canonical pathways. The top enriched GO included pathways involved in T cell differentiation in the thymus, negative regulation of cysteine-type endopeptidase activity, toll-like receptor, and endothelial cell differentiations (Fig. 3D).

Protein-Protein Interaction Network Construction and Hub Gene Analysis

Protein-protein interaction network analysis is a reliable method in understanding the biological responses in health and disease. Herein, protein interactions of the identified 72 markers were analyzed using the STRING database. A total of 46 nodes and 124 edges were revealed with combined scores >0.4 and visualized using Cytoscape software (Fig. 4A). The top 10 genes discovered using the CytoHubba plugin based on the MCC Algorithm are presented in Fig. 4B. The MCODE plugin identified two densely connected modules, including 14 markers (Fig. 4C). The final 9 genes panel was obtained by intersecting the 14 markers identified in two densely connected modules with the top 10 hub genes discovered by the MCC algorithm. PCA analysis (Fig. 4D) using the 9 genes panel separated DKD well from DM with a small overlap, which was also reasonable as DKD is developed from DM. Further, heatmap analysis showed that the final genes panel could well distinguish DKD from DM in our cohorts (Fig. 4E). Boxplot of 9 genes panel (CD28, CD44, PTEN, CXCR4, ITGAM, ITGB2, MYD88, CTNNB1, VCAM1) grouped by samples type (Additional file: Fig S1). Correlation analysis was conducted between the 9 individual genes and the clinical detection indicator eGFR. Data revealed that 8 genes were significantly negatively correlated with eGFR (Fig. 5).

Discussion

This study was focused on the discovery of hydroxymethylation-based biomarkers in plasma cfDNA that may facilitate the diagnostic of DKD. The availability of the blood samples and the convenience of cfDNA isolation makes it possible to monitor the occurrence and progression of the diseases with our technique. Thus, it may contribute to the personalized treatment and management of the disease.

Our data proved that the 5hmC signals were enriched in the promoter, exons, UTR, and TTS regions. Significant differences were found between DM and DKD for each genomic feature type, which suggested the key roles of these regions in the regulation of gene expression through 5hmC[14]. A number of DMGs derived from plasma cfDNA was detected by differential analysis. Alignment with the transcriptional sequencing data from corresponding glomerular and tubular tissue further identified 72 overlapping markers. GO analysis of the 72 markers suggested the enrichment of immune cell differentiation and proliferation, endothelial cell differentiation, and toll-like receptor-related signaling pathways. Previous studies[24, 25] have shown that activation of innate immune cells and resident kidney cells can trigger kidney inflammation, thereby promoting the pathogenesis and progression of DKD. Toll-like receptors detect endogenous risk-related molecular patterns produced during diabetes and induce inflammatory

responses in the sterile tubules through the NF- κ B signaling pathway. Therefore, anti-inflammatory therapies may have renal protection for DKD.

Finally, the results of the MCODE plugin and MCC algorithm were intersected to obtain the final hub genes. These hub genes exhibited good clustering between DM and DKD. Notably, 8 hub genes (CTNNB1, VCAM1, MYD88, CXCR4, ITGAM, CD28, ITGB2, CD44) were identified to have a significant negative correlation with eGFR. To our knowledge, β -Catenin (CTNNB1) plays key role in cell adhesion and regulates Wnt-mediated transcription. In the nucleus, CTNNB1 binds to lymphocyte enhancer factor-1/T cell factor transcription factors to mediate gene transcription. In addition, genome-wide association studies indicated a correlation between transcription factor 7 like 2 (TCF7L2) polymorphism and the development of diabetes and DKD[26, 27]. Previous research[28] revealed increased expression of Wnt/ β -catenin pathway transcripts and proteins in glomeruli and podocytes of patients and mouse models of DKD. Up-regulation of the Wnt/ β -catenin pathway in DKD may not only promote podocyte survival but also lead to cell detachment and podocyte loss. However, it's also been reported that down-regulation of the Wnt/ β -catenin pathway in podocytes might be important for terminal differentiation; however, it enhances apoptosis susceptibility. Therefore, we believe that balanced CTNNB1 expression is critical for glomerular filtration barrier maintenance in DKD. Additionally, we also found other immune cells surface marker genes (ITGAM, ITGB2, CD28, CD44), which lack functional studies with only appearing in the microarray information[29–31]. Meanwhile, in the correlation analysis, we found that the level of the epigenetic marker VCAM1 (Vascular Cell Adhesion Molecule 1) in the plasma cfDNA of DKD patients was highly negatively correlated with the level of eGFR, which was in consistent with the previous report[32]that renal filtration function gradually declines with the level of plasma VCAM1 increases. VCAM1 (Vascular Cell Adhesion Molecule 1) mainly encodes a cell surface sialoglycoprotein expressed by cytokine-activated endothelium and mediates leukocyte-endothelial cell adhesion and signal transduction that has been extensive-studied in cardiovascular diseases. The high expression of VCAM1 in renal tubular cells may promote their interactions with immune cells to affect infiltration, thereby promoting fibrosis[33]. More importantly, it has also been reported that the lipid-lowering drug Sitagliptin can maintain mitochondrial homeostasis through the SDF-1 α /CXCR4 pathway and improve DKD renal tubular injury[34]. 1,25(OH) $_2$ D $_3$ [35] inhibits inflammatory cell infiltration and tubulointerstitial fibrosis by downregulating the TLR4-MyD88-NF- κ B pathway and thus exhibits protective effects on DKD. Taking together, the evidence in published literature supported the reliability of our identified hub genes in terms of their roles in the occurrence and progression of DKD, and these hub genes may serve as potential drug targets for DKD.

Yet, there are limitations to our study. Firstly, the number of DKD patients is relatively small. Larger scale, independent cohorts will be necessary to refine and validate the genes panel. Secondly, although previously related studies have explored part of the functions of these hub genes, it still lacks the epigenetic mechanisms of the regulation of DKD by these DMGs. Finally, although cfDNA in the circulating blood is known to carry genetic and epigenetic information from cells of different origins, which is the fundamental basis of the cfDNA research, it's unlikely to estimate the precise proportion of

cfDNA derived from the lesion site at this moment. Future investigations regarding tissue-specific 5hmC will help solve this issue and provide novel insights into the epigenetic contribution to DKD.

Conclusions

In summary, the 5hmC-Seal assay was successfully applied in the cfDNA samples from a cohort of DM patients with or without DKD. Altered 5hmC markers were found in association with DKD, and that may contribute to gene regulation. Further data analysis indicated that cfDNA 5hmC signatures profiling might serve as a non-invasive epigenetic tool to distinguish DKD and to monitor DKD progression.

Abbreviations

DM: diabetes mellitus

DKD: diabetic kidney disease

eGFR: estimated glomerular filtration rate

cfDNA: cell-free DNA

5-hmC: 5-hydroxymethylcytosine

DEGs: differentially expressed genes

DMGs: differentially modified 5hmC marker genes

ESRD: end-stage renal disease

ADA: American Diabetes Association

UACR: urinary albumin-to-creatinine ratio

PPI: protein-protein interaction

PCA: principal component analysis

STRING: search tool for the retrieval of interacting genes

TSS: transcription start sites

TTS: transcription termination sites

Declarations

Ethics approval and consent to participate

The present study was conducted according to the guidelines of the Helsinki Declaration and was approved by the Ethics Committee of Peking University Third Hospital.

Consent for publication

Not applicable.

Availability of data and materials

All data generated or analyzed during this study are included within the article and its additional files.

Competing interests

The authors declare that they have no conflict of interest.

Funding

National Natural Science Foundation of China–Xinjiang Joint Fund (U1403322) for Excellent Young Scholars to L. L. National Natural Science Foundation of China (No. 81974506, 81673486, 81373405 and 30901803) to L. T.

Author contributions

J-LC, S-HB and YH conceived and designed the experiments. H-YC participated in library construct and 5hmC sequence with the help of R-YM and L-WD. J-LC analyzed data with help from H-YC, X-YW and M-ZZ. S-HB, YH, LZ, Z-QY, YM, GB and MD recruited patients, collected blood and organized clinical information. J-LC, R-YM and YH wrote the manuscript with input and comments from LC, H-YC and X-YW. All authors read and approved the final manuscript.

Acknowledgments

We would like to acknowledge the essential contributions of all staffs and students who participated in this work.

References

1. Thomas MC, Brownlee M, Susztak K, Sharma K, Jandeleit-Dahm KA, Zoungas S, Rossing P, Groop PH, Cooper ME: **Diabetic kidney disease.** *Nat Rev Dis Primers* 2015, **1**:15018.
2. Umanath K, Lewis JB: **Update on Diabetic Nephropathy: Core Curriculum 2018.** *Am J Kidney Dis* 2018, **71**:884-895.
3. Hanouneh M, Echouffo Tcheugui JB, Jaar BG: **Recent advances in diabetic kidney disease.** *BMC Med* 2021, **19**:180.
4. Perkins BA, Ficociello LH, Silva KH, Finkelstein DM, Warram JH, Krolewski AS: **Regression of microalbuminuria in type 1 diabetes.** *N Engl J Med* 2003, **348**:2285-2293.

5. Thomas MC, Macisaac RJ, Jerums G, Weekes A, Moran J, Shaw JE, Atkins RC: **Nonalbuminuric renal impairment in type 2 diabetic patients and in the general population (national evaluation of the frequency of renal impairment co-existing with NIDDM [NEFRON] 11).** *Diabetes Care* 2009, **32**:1497-1502.
6. Kato M, Natarajan R: **Epigenetics and epigenomics in diabetic kidney disease and metabolic memory.** *Nat Rev Nephrol* 2019, **15**:327-345.
7. Yang Y, Zeng C, Lu X, Song Y, Nie J, Ran R, Zhang Z, He C, Zhang W, Liu SM: **5-Hydroxymethylcytosines in Circulating Cell-Free DNA Reveal Vascular Complications of Type 2 Diabetes.** *Clin Chem* 2019, **65**:1414-1425.
8. Wang J, Tang J, Lai M, Zhang H: **5-Hydroxymethylcytosine and disease.** *Mutat Res Rev Mutat Res* 2014, **762**:167-175.
9. Song CX, Yin S, Ma L, Wheeler A, Chen Y, Zhang Y, Liu B, Xiong J, Zhang W, Hu J, et al: **5-Hydroxymethylcytosine signatures in cell-free DNA provide information about tumor types and stages.** *Cell Res* 2017, **27**:1231-1242.
10. Li W, Zhang X, Lu X, You L, Song Y, Luo Z, Zhang J, Nie J, Zheng W, Xu D, et al: **5-Hydroxymethylcytosine signatures in circulating cell-free DNA as diagnostic biomarkers for human cancers.** *Cell Res* 2017, **27**:1243-1257.
11. Han L, Chen C, Lu X, Song Y, Zhang Z, Zeng C, Chiu R, Li L, Xu M, He C, et al: **Alterations of 5-hydroxymethylcytosines in circulating cell-free DNA reflect retinopathy in type 2 diabetes.** *Genomics* 2021, **113**:79-87.
12. **2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes-2020.** *Diabetes Care* 2020, **43**:S14-s31.
13. **KDIGO 2020 Clinical Practice Guideline for Diabetes Management in Chronic Kidney Disease.** *Kidney Int* 2020, **98**:S1-s115.
14. Song CX, Szulwach KE, Fu Y, Dai Q, Yi C, Li X, Li Y, Chen CH, Zhang W, Jian X, et al: **Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine.** *Nat Biotechnol* 2011, **29**:68-72.
15. Chen HY, Zhang WL, Zhang L, Yang P, Li F, Yang ZR, Wang J, Pang M, Hong Y, Yan C, et al: **5-Hydroxymethylcytosine profiles of cfDNA are highly predictive of R-CHOP treatment response in diffuse large B cell lymphoma patients.** *Clin Epigenetics* 2021, **13**:33.
16. Langmead B, Salzberg SL: **Fast gapped-read alignment with Bowtie 2.** *Nat Methods* 2012, **9**:357-359.
17. Quinlan AR, Hall IM: **BEDTools: a flexible suite of utilities for comparing genomic features.** *Bioinformatics* 2010, **26**:841-842.
18. Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS: **Model-based analysis of ChIP-Seq (MACS).** *Genome Biol* 2008, **9**:R137.
19. Yu G, Wang LG, Han Y, He QY: **clusterProfiler: an R package for comparing biological themes among gene clusters.** *Omics* 2012, **16**:284-287.

20. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, et al: **STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets.** *Nucleic Acids Res* 2019, **47**:D607-d613.
21. Bandettini WP, Kellman P, Mancini C, Booker OJ, Vasu S, Leung SW, Wilson JR, Shanbhag SM, Chen MY, Arai AE: **MultiContrast Delayed Enhancement (MCODE) improves detection of subendocardial myocardial infarction by late gadolinium enhancement cardiovascular magnetic resonance: a clinical validation study.** *J Cardiovasc Magn Reson* 2012, **14**:83.
22. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY: **cytoHubba: identifying hub objects and sub-networks from complex interactome.** *BMC Syst Biol* 2014, **8 Suppl 4**:S11.
23. Zhang J, Han X, Gao C, Xing Y, Qi Z, Liu R, Wang Y, Zhang X, Yang YG, Li X, et al: **5-Hydroxymethylome in Circulating Cell-free DNA as A Potential Biomarker for Non-small-cell Lung Cancer.** *Genomics Proteomics Bioinformatics* 2018, **16**:187-199.
24. Tang SCW, Yiu WH: **Innate immunity in diabetic kidney disease.** *Nat Rev Nephrol* 2020, **16**:206-222.
25. Pérez-Morales RE, Del Pino MD, Valdivielso JM, Ortiz A, Mora-Fernández C, Navarro-González JF: **Inflammation in Diabetic Kidney Disease.** *Nephron* 2019, **143**:12-16.
26. Morgan MF, Salam RF, Rady NH, Alnaggar A, Ammar SH, Ghanem NS: **The Association of Transcription Factor 7 like 2 Gene Polymorphism with Diabetic Nephropathy in Patients with Type 2 Diabetes Mellitus.** *Curr Diabetes Rev* 2020, **16**:370-375.
27. Araoka T, Abe H, Tominaga T, Mima A, Matsubara T, Murakami T, Kishi S, Nagai K, Doi T: **Transcription factor 7-like 2 (TCF7L2) regulates activin receptor-like kinase 1 (ALK1)/Smad1 pathway for development of diabetic nephropathy.** *Mol Cells* 2010, **30**:209-218.
28. Kato H, Gruenwald A, Suh JH, Miner JH, Barisoni-Thomas L, Taketo MM, Faul C, Millar SE, Holzman LB, Susztak K: **Wnt/ β -catenin pathway in podocytes integrates cell adhesion, differentiation, and survival.** *J Biol Chem* 2011, **286**:26003-26015.
29. Mastek K, Adamiec R: **Neutrophil surface expression of CD11b and CD62L in diabetic microangiopathy.** *Acta Diabetol* 2008, **45**:183-190.
30. Geng XD, Wang WW, Feng Z, Liu R, Cheng XL, Shen WJ, Dong ZY, Cai GY, Chen XM, Hong Q, Wu D: **Identification of key genes and pathways in diabetic nephropathy by bioinformatics analysis.** *J Diabetes Investig* 2019, **10**:972-984.
31. Li Y, Jin L, Yan J, Zhang H, Zhang R, Hu C: **CD28 Genetic Variants Increase Susceptibility to Diabetic Kidney Disease in Chinese Patients with Type 2 Diabetes: A Cross-Sectional Case Control Study.** *Mediators Inflamm* 2021, **2021**:5521050.
32. Liu JJ, Yeoh LY, Sum CF, Tavintharan S, Ng XW, Liu S, Lee SB, Tang WE, Lim SC: **Vascular cell adhesion molecule-1, but not intercellular adhesion molecule-1, is associated with diabetic kidney disease in Asians with type 2 diabetes.** *J Diabetes Complications* 2015, **29**:707-712.
33. Jia Y, Xu H, Yu Q, Tan L, Xiong Z: **Identification and verification of vascular cell adhesion protein 1 as an immune-related hub gene associated with the tubulointerstitial injury in diabetic kidney disease.**

34. Zhang Q, He L, Dong Y, Fei Y, Wen J, Li X, Guan J, Liu F, Zhou T, Li Z, et al: **Sitagliptin ameliorates renal tubular injury in diabetic kidney disease via STAT3-dependent mitochondrial homeostasis through SDF-1 α /CXCR4 pathway.** *Faseb j* 2020, **34**:7500-7519.
35. Liu P, Li F, Xu X, Li S, Dong X, Chen L, Bai B, Wang Y, Qiu M, Dong Y: **1,25(OH)(2)D(3) provides protection against diabetic kidney disease by downregulating the TLR4-MyD88-NF- κ B pathway.** *Exp Mol Pathol* 2020, **114**:104434.

Table

Table 1 Characteristics of the study subjects

Characteristics	Control(n=13)	DM(n=14)	DKD(n=17)	P-Value
Age (mean (SD))	43.4 (7.2)	47.9 (9.9)	51.1 (16.7)	0.257
Gender(male/female)	6/7	9/5	9/8	0.902
Glucose(mean(SD),mmol/L)	4.2(1.1)	8.9(1.6)	8.3(1.8)	<0.001
TC(mean(SD),mmol/L)	4.4(0.7)	4.4(0.7)	4.5(0.5)	0.753
TG(mean(SD),mmol/L)	1.3(0.5)	1.6(0.4)	1.7(1.0)	0.190
LDL(mean(SD),mmol/L)	2.85(0.4)	2.5(0.7)	2.5(0.5)	0.212
HDL(mean(SD),mmol/L)	1.2(0.4)	1.3(0.4)	1.3(0.2)	0.469
HbA1c (mean(SD),%)	-	8.3(1.1)	6.8(1.1)	<0.001
Creatinine (mean(SD), μ mol/L)	71.6(9.0)	71.3 (12.8)	112.9(44.4)	<0.001
Serum Urea (mean(SD), mmol/L)	5.4(1.4)	5.8(1.2)	7.8(2.8)	<0.05
CysC (mean(SD), mg/L)	-	0.7(0.1)	1.4(0.6)	<0.001
eGFR(mean(SD), mL/min/1.73m ²)	-	98.4(6.8)	53.9(25.0)	<0.001
UACR(mean(SD), g/g)	-	-	2.5 (1.6)	-
24-h urine protein (mean(SD), g/d)	-	-	1.7(0.4)	-

Data are presented as mean $\pm$ SD (standard deviation)

TC: total cholesterol; TG: triglycerides; LDL: lowdensity lipoprotein cholesterol; HDL: high density lipoprotein; HbA1c: glycated hemoglobin; CysC: cystatin-C; eGFR: estimated glomerular filtration rate; UACR: urinary albumin to creatinine ratio; DM: diabetes mellitus; DKD: diabetic kidney disease

Figures

Fig.1

Figure 1

Overview of study design.

Figure 2

Characteristics of 5hmC distribution in plasma cfDNA of DKD patients. (A) The profiled 5hmC-Seal data in all samples cfDNA are enriched in gene bodies and depleted in the flanking regions. (B) Number of 5hmC peaks detected per million reads in Control, DM, and DKD cohorts. Each dot depicts an individual

sample. (C) Genome-wide 5hmC distribution in different genomic features grouped by 3 groups (Control vs. DM vs. DKD). (D) Volcano plot. Significantly altered DMGs ($|\log_2FC| > 0.5$, $p\text{-value} < 0.05$) are highlighted in red (up) or green (down) using the DKD vs DM cfDNA samples. Black dots represent the genes that are not differentially expressed. (E) Mean $\log_2$ Foldchange value of 5168 DMGs across different genomic features. (F) Pathways enriched in the upregulated marker genes with modified 5hmC between patients with and without DKD are shown. (G) Pathways enriched in the downregulated marker genes with modified 5hmC between patients. (H) Heatmap of top 200 DMGs with sample type, age, and sex information labeled. Unsupervised hierarchical clustering was performed across genes and samples. RPM: Reads of exon model per Million mapped reads, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.001$

Figure 3

Performance of the DMGs in cfDNA for patients (DM vs. DKD). (A) An upset diagram of 62 intersected genes was found in upregulated genes via taking the intersection of DMGs from 5hmC-Seq and DEGs from GSE30528 and GSE30529. (B) 10 intersected and downregulated genes among our cohort and GSE30528 and GSE30529. (C) IGV genome browser snapshot of CTNNB1 locus showing the increased

5hmC signal intensity in DKD samples compared to Control and DM. (D) GO enrichment analysis and function exploration of 72 DMGs using Cytoscape software. (*p<0.05, **p<0.01).

Figure 4

The final 9 genes panel could well distinguish DKD from DM. (A) Based on database STRING and Cytoscape software, PPI networks of 72 DMGs were constructed. The darker the color of the node, the

greater the degree value. (B) Hub genes selection and analysis performed by cytohubba. The score of hub genes was based on the MCC Algorithm. (C) Modular analyses of PPI found two key modules. (D) PCA plots showing DM (orange) and DKD (red) cfDNA cohorts using 9 genes panel as features. (E) Heatmaps of 9 genes panel with sample type, age, and sex information labeled in our cohort. Unsupervised hierarchical clustering was performed across genes and samples.

Figure 5

Correlation analysis between the expression level of cfDNA derived 9 DMGs and the clinical parameters in DKD patients. The significant negative correlation could be found among the expression level of CD28, CD44, CTNNB1, CXCR4, ITGAM, ITGB2, MYD88, VCAM1 with eGFR.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [FigS1.jpg](#)