Supplementary Materials

Functional methylome analysis of human diabetic kidney disease

Jihwan Park[1, †], Yuting Guan[1, †], Xin Sheng[1], Caroline Gluck[1], Matthew J. Seasock[1], Ari Hakimi[2], Chengxiang Qiu[1], James Pullman[3], Amit Verma[4], Hongzhe Li[5], Matthew Palmer[6] and Katalin Susztak[1,*]

[1]Department of Medicine, Renal Electrolyte and Hypertension Division, Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA, 19104

[2]Department of Surgery, Urology Service, Memorial Sloan Kettering Medical Center, New York, NY,

[3]Department of Pathology, Montefiore Medical Center, Bronx, NY,

[4]Department of Oncology and Developmental Biology, Albert Einstein College of Medicine, Bronx, NY, USA

[5]Department of Epidemiology and Biostatistics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA, 19104

[6]Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA, 19104

†[7]These authors contributed equally to this work.

*Corresponding author: Katalin Susztak

12-123 Smilow Translational Research Center

3400 Civic Center Blvd

Philadelphia, PA 19104

1(215)-898-2009

ksusztak@pennmedicine.upenn.edu

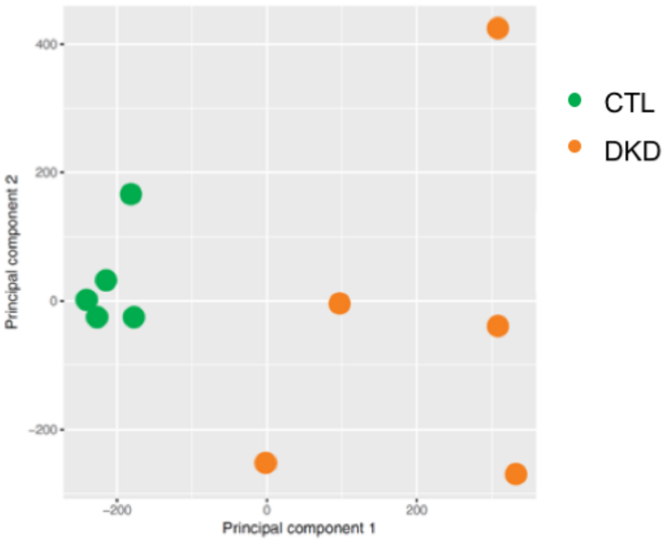
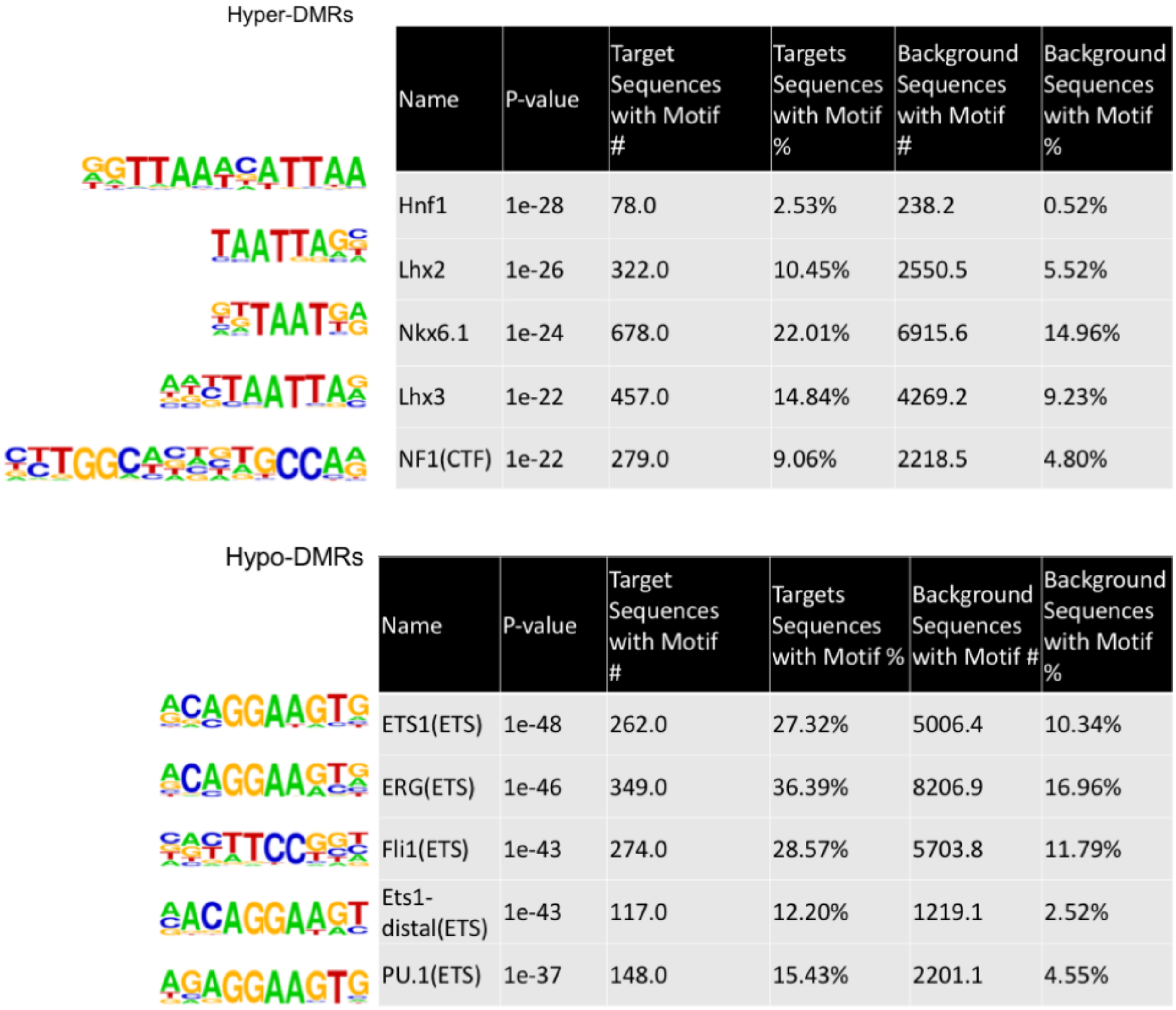


Figure S1. **PCA plot for the WGBS data of microdissected 10 human kidney tubules.** X-axis and Y-axis represent PC1 and PC2, respectively.



Hyper-DMRs

Name	P-value	Target Sequences with Motif #	Targets Sequences with Motif %	Background Sequences with Motif #	Background Sequences with Motif %
Hnf1	1e-28	78.0	2.53%	238.2	0.52%
Lhx2	1e-26	322.0	10.45%	2550.5	5.52%
Nkx6.1	1e-24	678.0	22.01%	6915.6	14.96%
Lhx3	1e-22	457.0	14.84%	4269.2	9.23%
NF1(CTF)	1e-22	279.0	9.06%	2218.5	4.80%

Hypo-DMRs

Name	P-value	Target Sequences with Motif #	Targets Sequences with Motif %	Background Sequences with Motif #	Background Sequences with Motif %
ETS1(ETS)	1e-48	262.0	27.32%	5006.4	10.34%
ERG(ETS)	1e-46	349.0	36.39%	8206.9	16.96%
Fli1(ETS)	1e-43	274.0	28.57%	5703.8	11.79%
Ets1-distal(ETS)	1e-43	117.0	12.20%	1219.1	2.52%
PU.1(ETS)	1e-37	148.0	15.43%	2201.1	4.55%

Figure S2. Enrichment analysis of transcription factor (TF) binding motifs in the DMRs using HOMER package. Top 5 TF binding motifs are selected based on p-values.

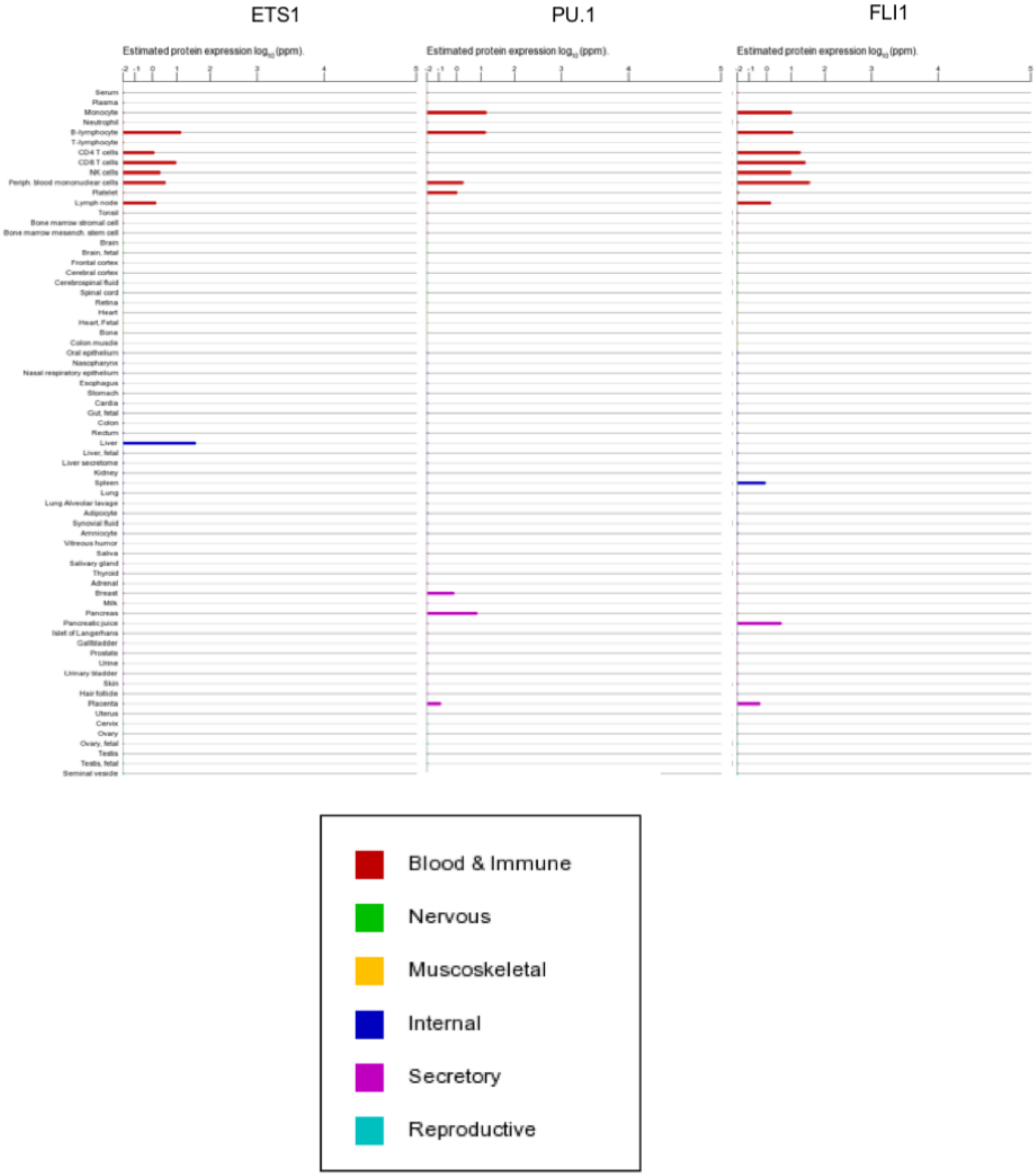
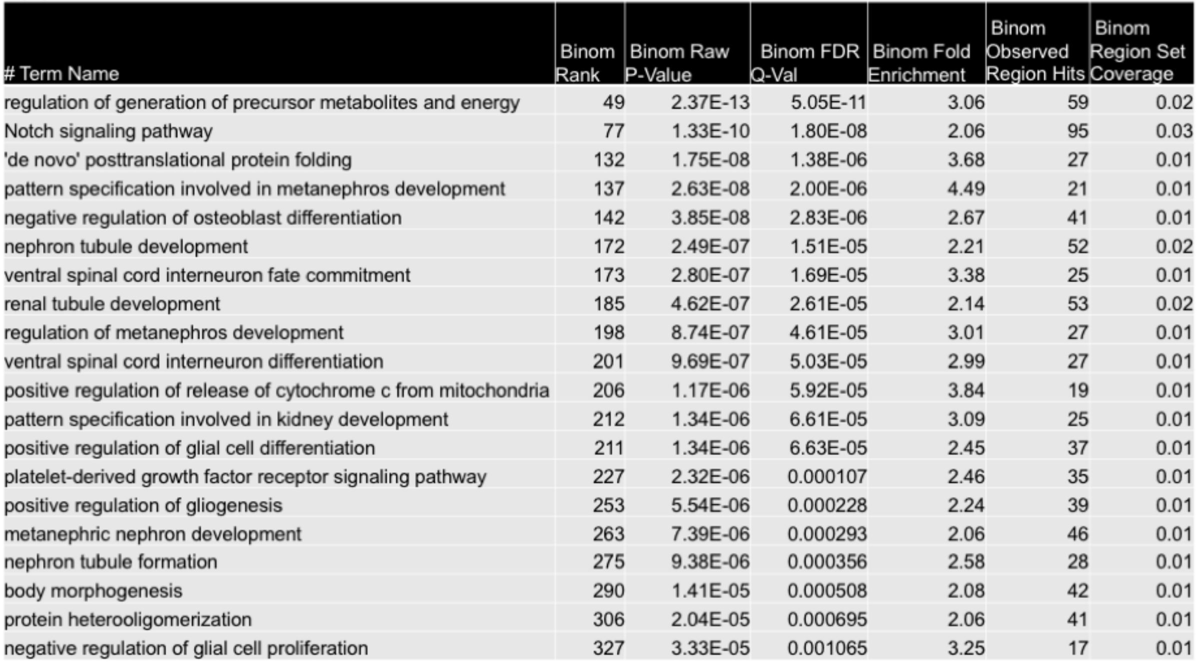


Figure S3. Estimated protein expression levels in various tissue and cell types (1). Blood cell types are marked red.

A

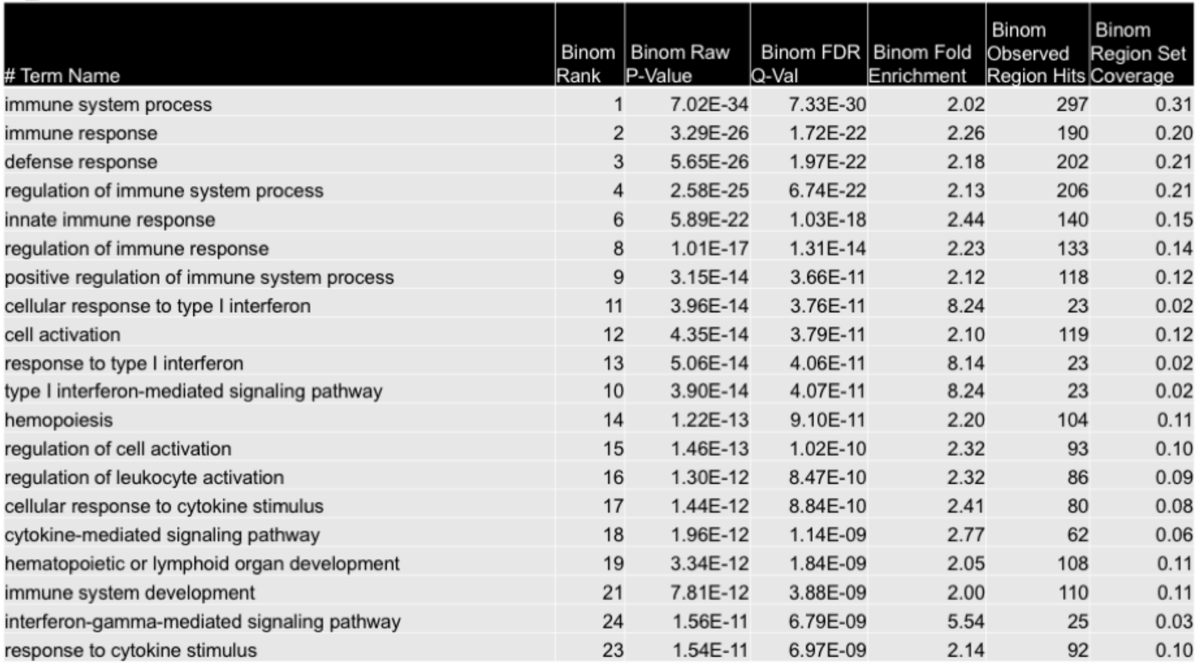
Hyper-DMRs



# Term Name	Binom Rank	Binom Raw P-Value	Binom FDR Q-Val	Binom Fold Enrichment	Binom Observed Region Hits	Binom Region Set Coverage
regulation of generation of precursor metabolites and energy	49	2.37E-13	5.05E-11	3.06	59	0.02
Notch signaling pathway	77	1.33E-10	1.80E-08	2.06	95	0.03
'de novo' posttranslational protein folding	132	1.75E-08	1.38E-06	3.68	27	0.01
pattern specification involved in metanephros development	137	2.63E-08	2.00E-06	4.49	21	0.01
negative regulation of osteoblast differentiation	142	3.85E-08	2.83E-06	2.67	41	0.01
nephron tubule development	172	2.49E-07	1.51E-05	2.21	52	0.02
ventral spinal cord interneuron fate commitment	173	2.80E-07	1.69E-05	3.38	25	0.01
renal tubule development	185	4.62E-07	2.61E-05	2.14	53	0.02
regulation of metanephros development	198	8.74E-07	4.61E-05	3.01	27	0.01
ventral spinal cord interneuron differentiation	201	9.69E-07	5.03E-05	2.99	27	0.01
positive regulation of release of cytochrome c from mitochondria	206	1.17E-06	5.92E-05	3.84	19	0.01
pattern specification involved in kidney development	212	1.34E-06	6.61E-05	3.09	25	0.01
positive regulation of glial cell differentiation	211	1.34E-06	6.63E-05	2.45	37	0.01
platelet-derived growth factor receptor signaling pathway	227	2.32E-06	0.000107	2.46	35	0.01
positive regulation of gliogenesis	253	5.54E-06	0.000228	2.24	39	0.01
metanephric nephron development	263	7.39E-06	0.000293	2.06	46	0.01
nephron tubule formation	275	9.38E-06	0.000356	2.58	28	0.01
body morphogenesis	290	1.41E-05	0.000508	2.08	42	0.01
protein heterooligomerization	306	2.04E-05	0.000695	2.06	41	0.01
negative regulation of glial cell proliferation	327	3.33E-05	0.001065	3.25	17	0.01

B

Hypo-DMRs



# Term Name	Binom Rank	Binom Raw P-Value	Binom FDR Q-Val	Binom Fold Enrichment	Binom Observed Region Hits	Binom Region Set Coverage
immune system process	1	7.02E-34	7.33E-30	2.02	297	0.31
immune response	2	3.29E-26	1.72E-22	2.26	190	0.20
defense response	3	5.65E-26	1.97E-22	2.18	202	0.21
regulation of immune system process	4	2.58E-25	6.74E-22	2.13	206	0.21
innate immune response	6	5.89E-22	1.03E-18	2.44	140	0.15
regulation of immune response	8	1.01E-17	1.31E-14	2.23	133	0.14
positive regulation of immune system process	9	3.15E-14	3.66E-11	2.12	118	0.12
cellular response to type I interferon	11	3.96E-14	3.76E-11	8.24	23	0.02
cell activation	12	4.35E-14	3.79E-11	2.10	119	0.12
response to type I interferon	13	5.06E-14	4.06E-11	8.14	23	0.02
type I interferon-mediated signaling pathway	10	3.90E-14	4.07E-11	8.24	23	0.02
hemopoiesis	14	1.22E-13	9.10E-11	2.20	104	0.11
regulation of cell activation	15	1.46E-13	1.02E-10	2.32	93	0.10
regulation of leukocyte activation	16	1.30E-12	8.47E-10	2.32	86	0.09
cellular response to cytokine stimulus	17	1.44E-12	8.84E-10	2.41	80	0.08
cytokine-mediated signaling pathway	18	1.96E-12	1.14E-09	2.77	62	0.06
hematopoietic or lymphoid organ development	19	3.34E-12	1.84E-09	2.05	108	0.11
immune system development	21	7.81E-12	3.88E-09	2.00	110	0.11
interferon-gamma-mediated signaling pathway	24	1.56E-11	6.79E-09	5.54	25	0.03
response to cytokine stimulus	23	1.54E-11	6.97E-09	2.14	92	0.10

Figure S4. GREAT functional annotation analysis of the hyper- (A) and hypo-DMRs (B).

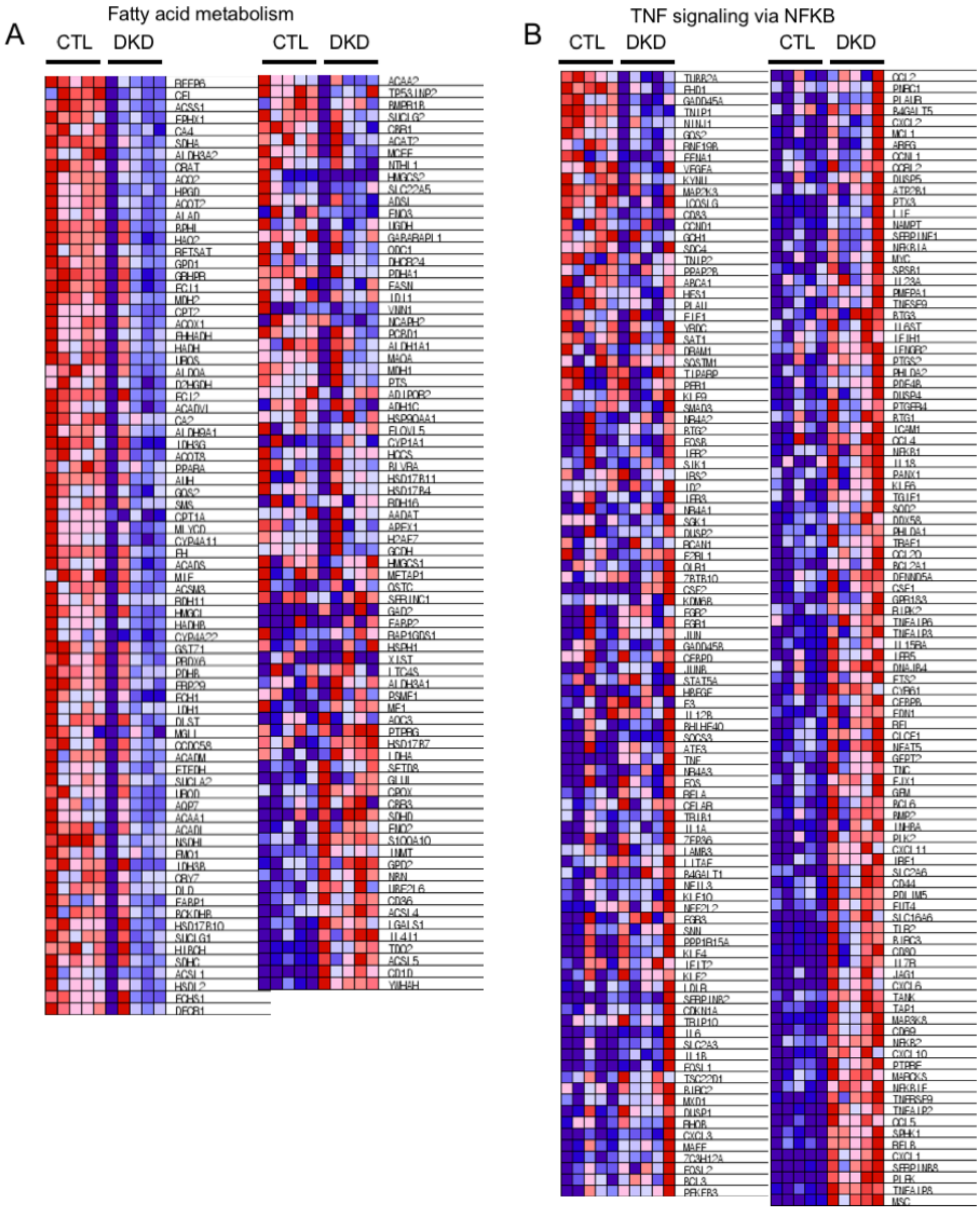
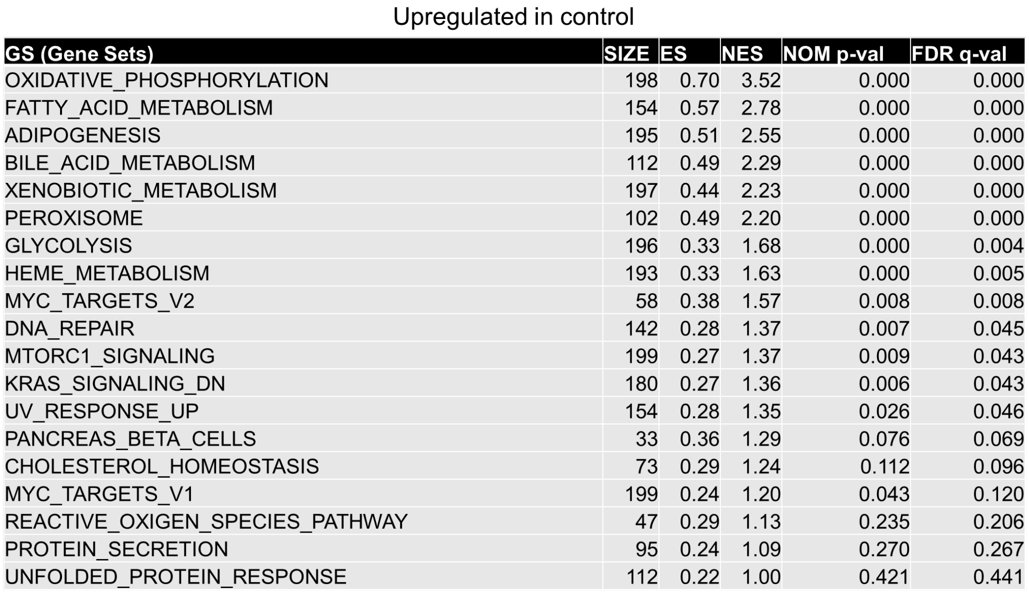
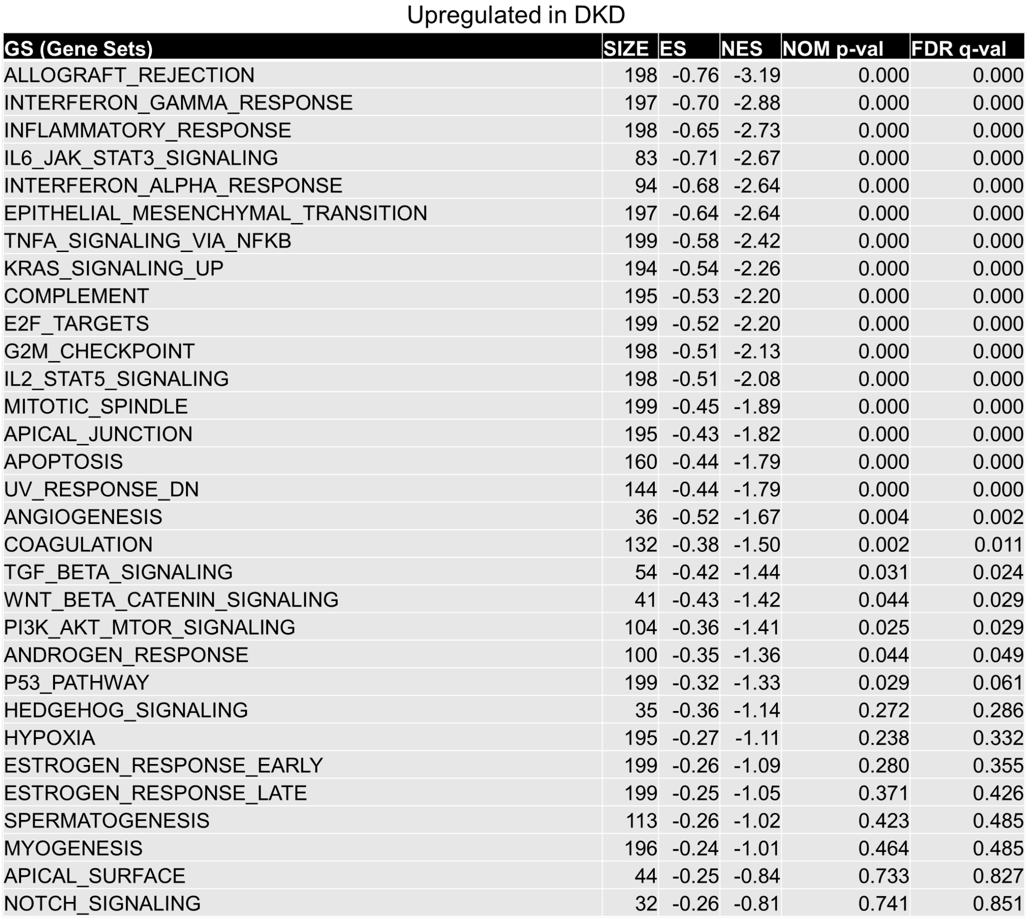


Figure S5. Selected gene set enrichment analysis (GSEA) results of genes that are increased and decreased in DKD. Color represents relative expression level of each gene from low (blue) to high (red) in DKD samples. (**A**) Genes associated with fatty acid metabolism. (**B**) Genes associated with *TNF* signaling via *NFKB*.



Upregulated in control

GS (Gene Sets)	SIZE	ES	NES	NOM p-val	FDR q-val
OXIDATIVE_PHOSPHORYLATION	198	0.70	3.52	0.000	0.000
FATTY_ACID_METABOLISM	154	0.57	2.78	0.000	0.000
ADIPOGENESIS	195	0.51	2.55	0.000	0.000
BILE_ACID_METABOLISM	112	0.49	2.29	0.000	0.000
XENOBIOTIC_METABOLISM	197	0.44	2.23	0.000	0.000
PEROXISOME	102	0.49	2.20	0.000	0.000
GLYCOLYSIS	196	0.33	1.68	0.000	0.004
HEME_METABOLISM	193	0.33	1.63	0.000	0.005
MYC_TARGETS_V2	58	0.38	1.57	0.008	0.008
DNA_REPAIR	142	0.28	1.37	0.007	0.045
MTORC1_SIGNALING	199	0.27	1.37	0.009	0.043
KRAS_SIGNALING_DN	180	0.27	1.36	0.006	0.043
UV_RESPONSE_UP	154	0.28	1.35	0.026	0.046
PANCREAS_BETA_CELLS	33	0.36	1.29	0.076	0.069
CHOLESTEROL_HOMEOSTASIS	73	0.29	1.24	0.112	0.096
MYC_TARGETS_V1	199	0.24	1.20	0.043	0.120
REACTIVE_OXIGEN_SPECIES_PATHWAY	47	0.29	1.13	0.235	0.206
PROTEIN_SECRETION	95	0.24	1.09	0.270	0.267
UNFOLDED_PROTEIN_RESPONSE	112	0.22	1.00	0.421	0.441



Upregulated in DKD

GS (Gene Sets)	SIZE	ES	NES	NOM p-val	FDR q-val
ALLOGRAFT_REJECTION	198	-0.76	-3.19	0.000	0.000
INTERFERON_GAMMA_RESPONSE	197	-0.70	-2.88	0.000	0.000
INFLAMMATORY_RESPONSE	198	-0.65	-2.73	0.000	0.000
IL6_JAK_STAT3_SIGNALING	83	-0.71	-2.67	0.000	0.000
INTERFERON_ALPHA_RESPONSE	94	-0.68	-2.64	0.000	0.000
EPITHELIAL_MESENCHYMAL_TRANSITION	197	-0.64	-2.64	0.000	0.000
TNFA_SIGNALING_VIA_NFKB	199	-0.58	-2.42	0.000	0.000
KRAS_SIGNALING_UP	194	-0.54	-2.26	0.000	0.000
COMPLEMENT	195	-0.53	-2.20	0.000	0.000
E2F_TARGETS	199	-0.52	-2.20	0.000	0.000
G2M_CHECKPOINT	198	-0.51	-2.13	0.000	0.000
IL2_STAT5_SIGNALING	198	-0.51	-2.08	0.000	0.000
MITOTIC_SPINDLE	199	-0.45	-1.89	0.000	0.000
APICAL_JUNCTION	195	-0.43	-1.82	0.000	0.000
APOPTOSIS	160	-0.44	-1.79	0.000	0.000
UV_RESPONSE_DN	144	-0.44	-1.79	0.000	0.000
ANGIOGENESIS	36	-0.52	-1.67	0.004	0.002
COAGULATION	132	-0.38	-1.50	0.002	0.011
TGF_BETA_SIGNALING	54	-0.42	-1.44	0.031	0.024
WNT_BETA_CATENIN_SIGNALING	41	-0.43	-1.42	0.044	0.029
PI3K_AKT_MTOR_SIGNALING	104	-0.36	-1.41	0.025	0.029
ANDROGEN_RESPONSE	100	-0.35	-1.36	0.044	0.049
P53_PATHWAY	199	-0.32	-1.33	0.029	0.061
HEDGEHOG_SIGNALING	35	-0.36	-1.14	0.272	0.286
HYPOXIA	195	-0.27	-1.11	0.238	0.332
ESTROGEN_RESPONSE_EARLY	199	-0.26	-1.09	0.280	0.355
ESTROGEN_RESPONSE_LATE	199	-0.25	-1.05	0.371	0.426
SPERMATOGENESIS	113	-0.26	-1.02	0.423	0.485
MYOGENESIS	196	-0.24	-1.01	0.464	0.485
APICAL_SURFACE	44	-0.25	-0.84	0.733	0.827
NOTCH_SIGNALING	32	-0.26	-0.81	0.741	0.851

Figure S6. Gene set enrichment analysis (GSEA) results of genes that are upregulated in control or in DKD .

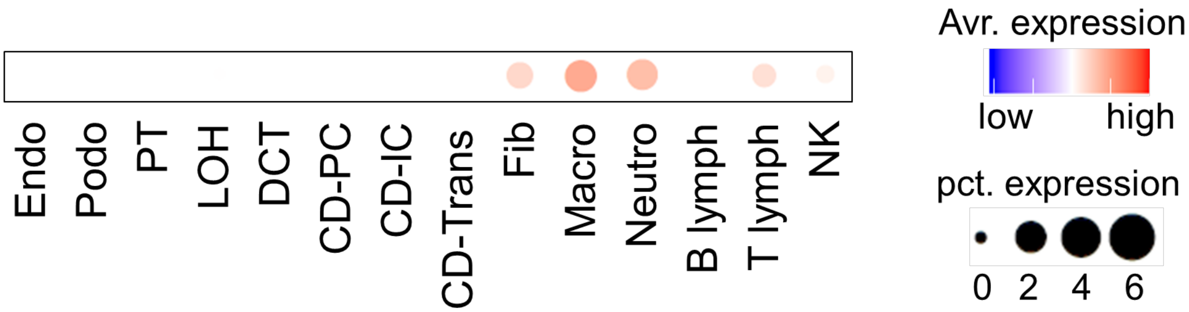


Figure S7. The expression of *Tnf* in mouse kidney cells as identified by single cell RNA sequencing.

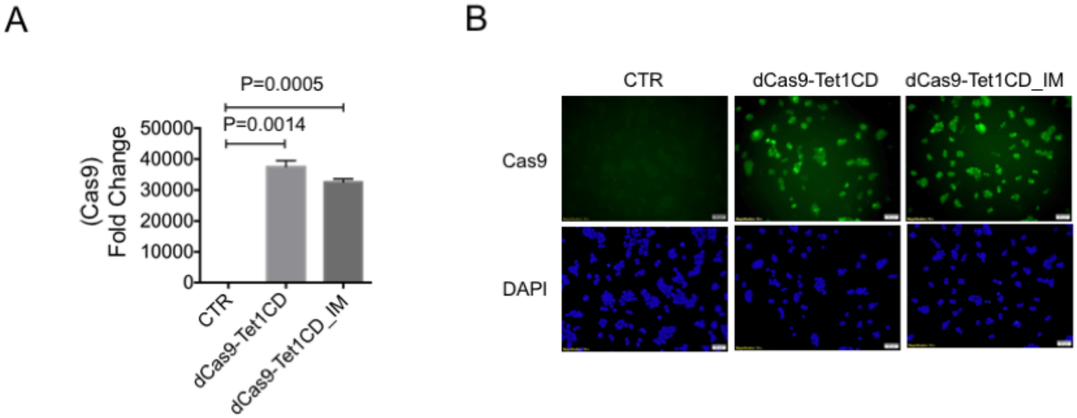


Figure S8. Generation of cell lines with dCas9-Tet1. (A) Relative transcript level of Cas9 in the control, dCas9-Tet1CD cell clones and dCas9-Tet1CD_IM cells. **(B)** Representative images of Cas9 immunostaining of control, dCas9-Tet1CD cells and dCas9-Tet1CD_IM cells. Data are presented as means ± SEM and analyzed by one-way ANOVA with Tukey's post hoc tests.

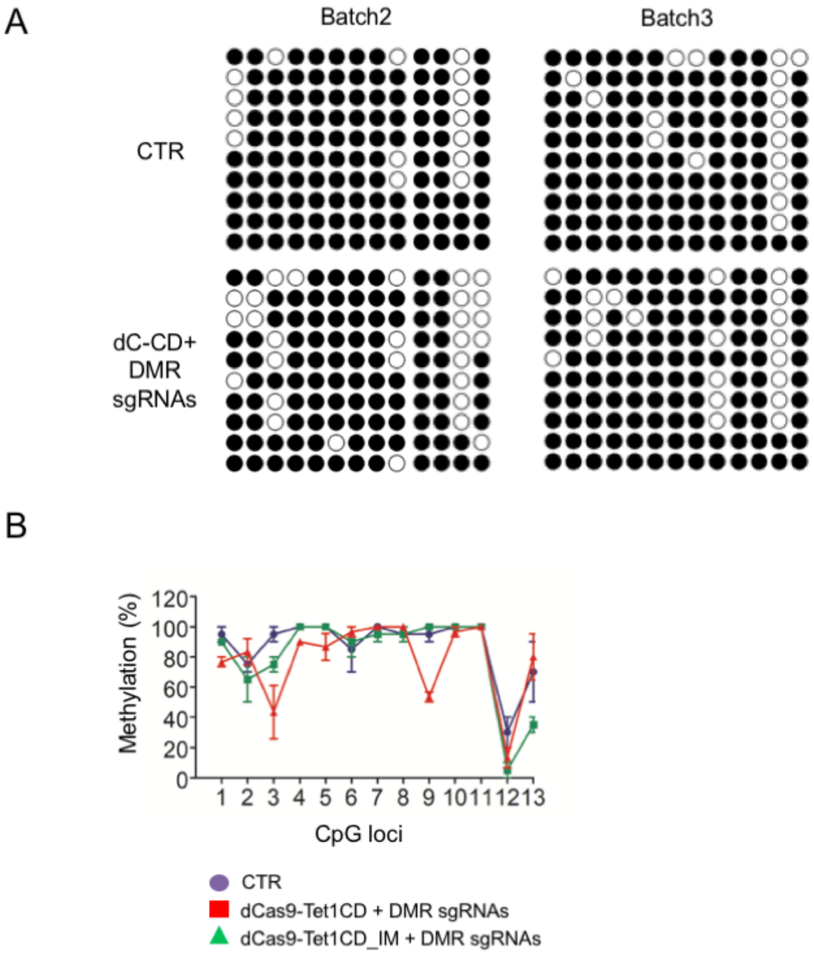


Figure S9. Bisulfite PCR sequencing. (A) Bisulfite sequencing analysis of TNF DMR in control and dCas9-Tet1CD cells transfected with DMR sgRNAs. Batch 2 and 3 were independent experiment. **(B)** Combined methylation level of 13 CpG loci in *TNF* DMR in cells. Data are presented as means ± SEM.

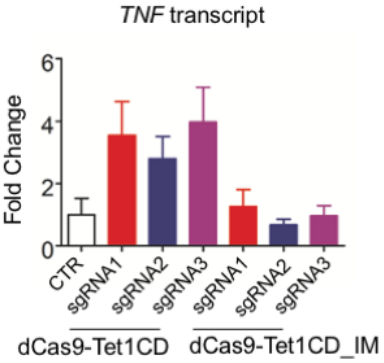


Figure S10. ***TNF*** **transcript level of dCas9-Tet1CD or control cells transfected with guide RNA.** Relative transcript level of *TNF* in Cas9 control (CTR), dCas9-Tet1CD cells or dCas9-Tet1CD_IM cells transfected with single sgRNA 1, 2, 3. Data represent means ± SE.

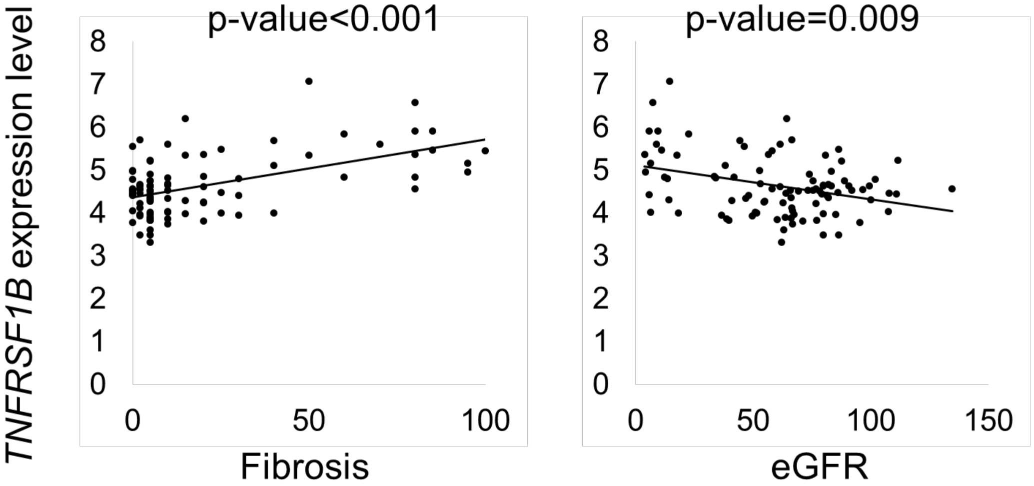


Figure S11. Gene expression results showing the expression levels of *TNFRSF1B* gene in human control and DKD samples. In human patient samples *TNFRSF1B* gene expression was associated with interstitial fibrosis and kidney function (eGFR). P-value was calculated using linear regression models adjusted for age, sex, race, diabetes and hypertension (2).

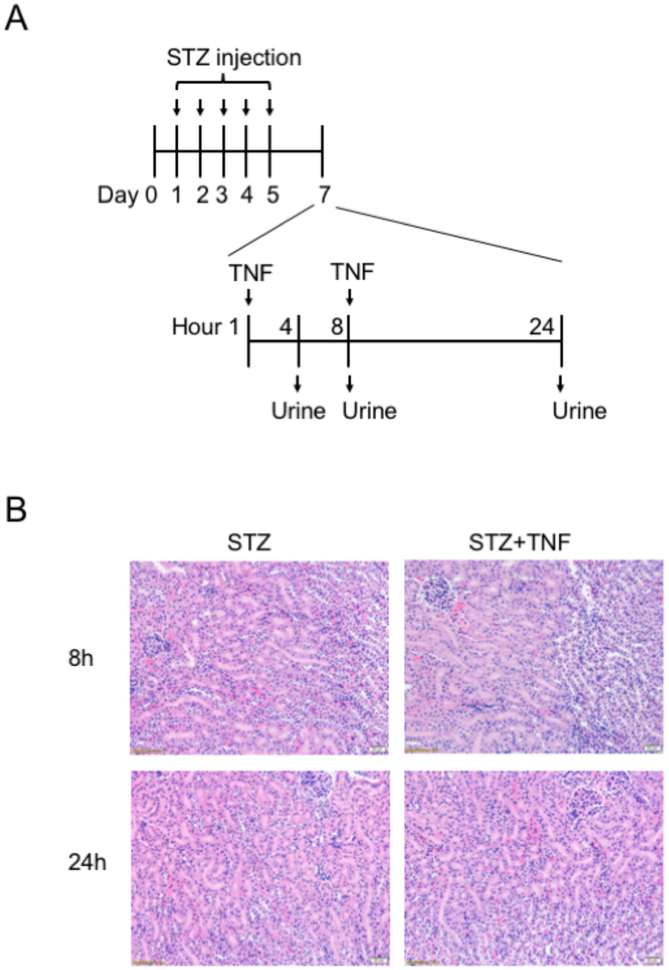
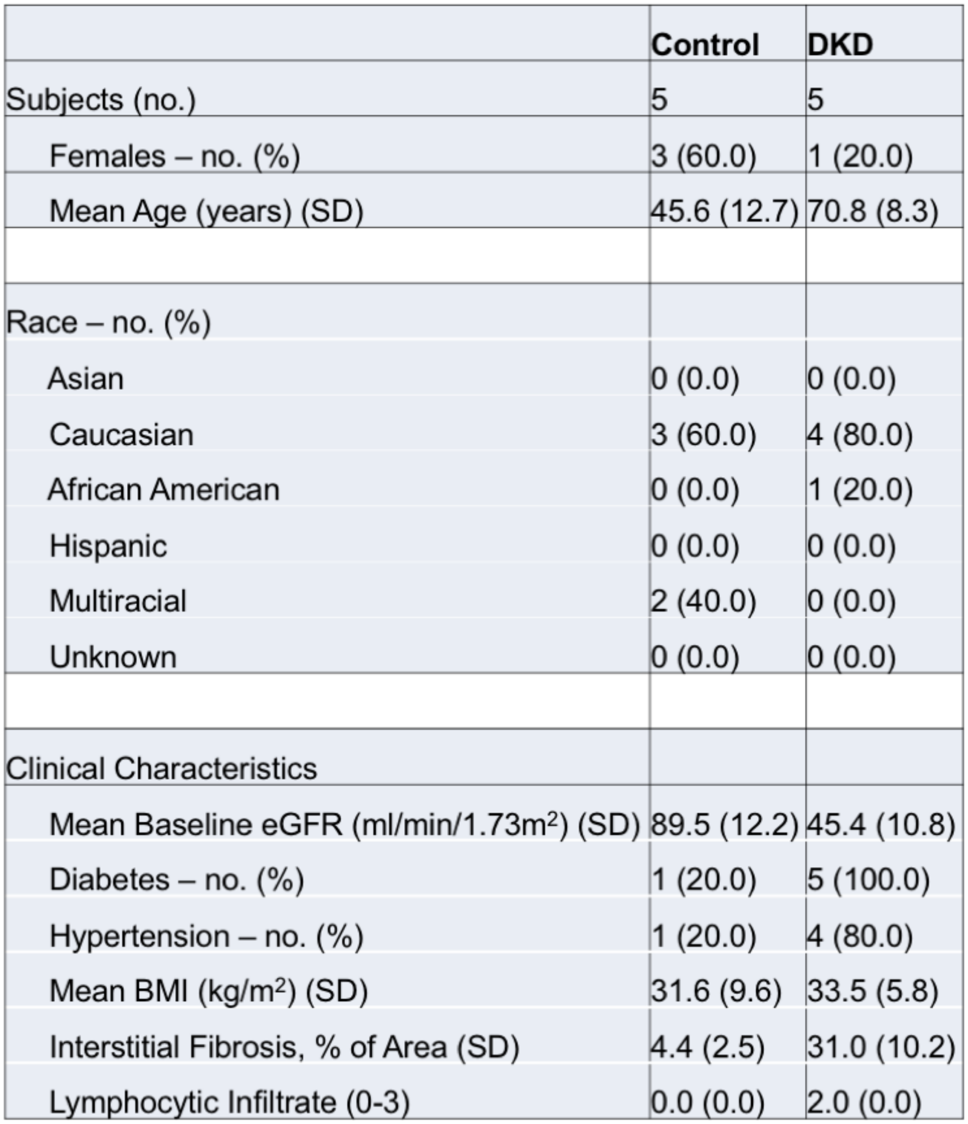
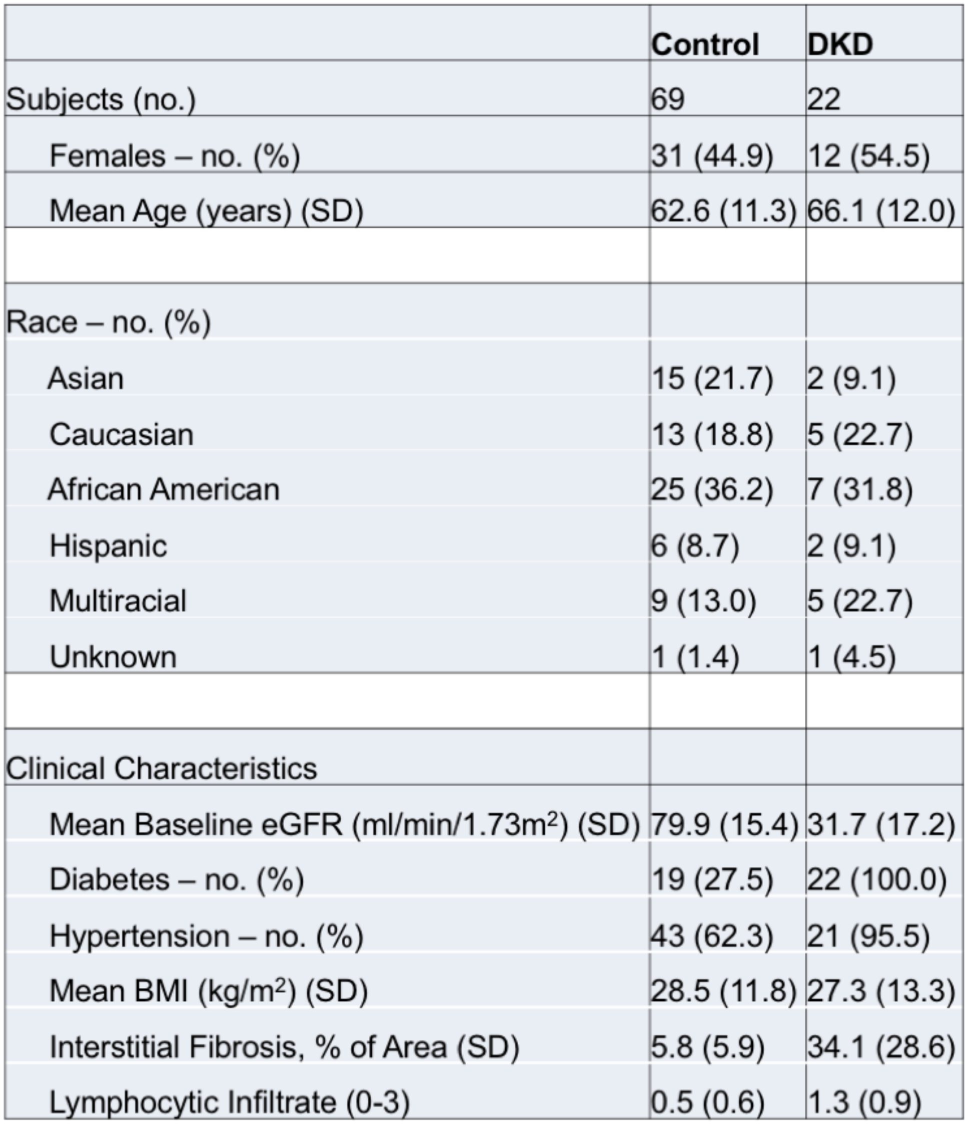


Figure S12. Representative kidney histological images of control, diabetic or mice treated with TNF. (A) Schematic diagram of STZ and TNF treatment and urine collection. **(B)** Representative images of H&E stained kidney sections from control STZ-treated and STZ/TNF-treated mice. Scale bars, 20 um.



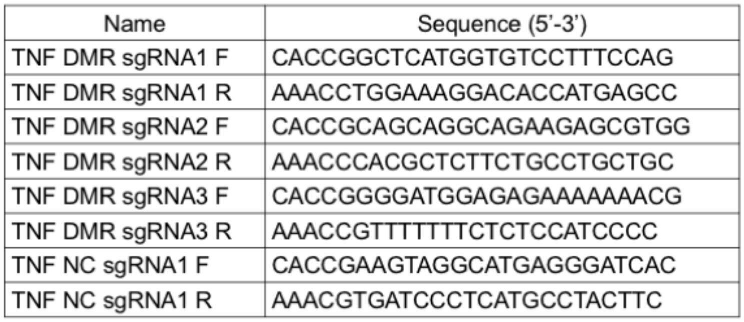
	Control	DKD
Subjects (no.)	5	5
Females – no. (%)	3 (60.0)	1 (20.0)
Mean Age (years) (SD)	45.6 (12.7)	70.8 (8.3)
Race – no. (%)		
Asian	0 (0.0)	0 (0.0)
Caucasian	3 (60.0)	4 (80.0)
African American	0 (0.0)	1 (20.0)
Hispanic	0 (0.0)	0 (0.0)
Multiracial	2 (40.0)	0 (0.0)
Unknown	0 (0.0)	0 (0.0)
Clinical Characteristics		
Mean Baseline eGFR (ml/min/1.73m^2) (SD)	89.5 (12.2)	45.4 (10.8)
Diabetes – no. (%)	1 (20.0)	5 (100.0)
Hypertension – no. (%)	1 (20.0)	4 (80.0)
Mean BMI (kg/m^2) (SD)	31.6 (9.6)	33.5 (5.8)
Interstitial Fibrosis, % of Area (SD)	4.4 (2.5)	31.0 (10.2)
Lymphocytic Infiltrate (0-3)	0.0 (0.0)	2.0 (0.0)

Table S1. Demographic and clinical characteristics of data sets. The values are the mean and standard deviations shown in parentheses (SD) or the absolute number with the percentage of patients shown in parentheses. The value for interstitial fibrosis is the percentage of the biopsy sample showing fibrosis, with 0% being no fibrosis and 100% being complete fibrosis. The value for lymphocytic infiltrate was scored by histology analysis of the biopsy samples (0-3). Abbreviations: eGFR, estimated glomerular filtration rate; BMI, body mass index.



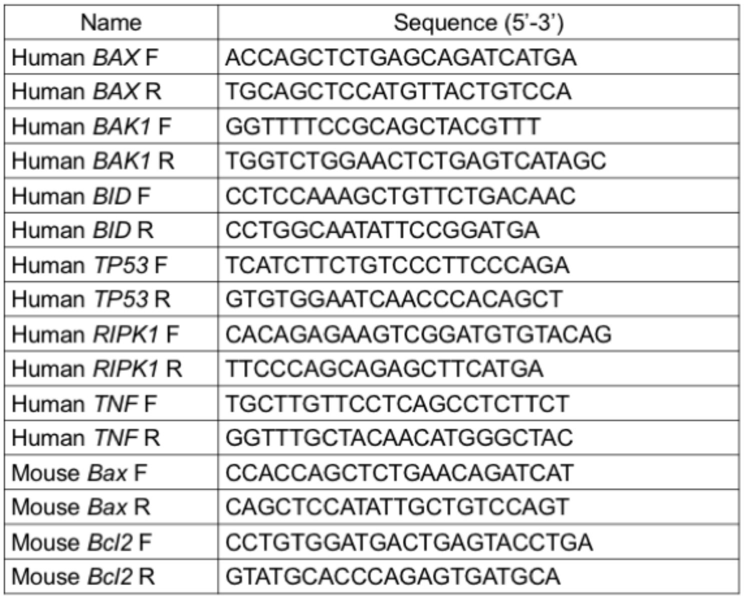
	Control	DKD
Subjects (no.)	69	22
Females – no. (%)	31 (44.9)	12 (54.5)
Mean Age (years) (SD)	62.6 (11.3)	66.1 (12.0)
Race – no. (%)		
Asian	15 (21.7)	2 (9.1)
Caucasian	13 (18.8)	5 (22.7)
African American	25 (36.2)	7 (31.8)
Hispanic	6 (8.7)	2 (9.1)
Multiracial	9 (13.0)	5 (22.7)
Unknown	1 (1.4)	1 (4.5)
Clinical Characteristics		
Mean Baseline eGFR (ml/min/1.73m^2) (SD)	79.9 (15.4)	31.7 (17.2)
Diabetes – no. (%)	19 (27.5)	22 (100.0)
Hypertension – no. (%)	43 (62.3)	21 (95.5)
Mean BMI (kg/m^2) (SD)	28.5 (11.8)	27.3 (13.3)
Interstitial Fibrosis, % of Area (SD)	5.8 (5.9)	34.1 (28.6)
Lymphocytic Infiltrate (0-3)	0.5 (0.6)	1.3 (0.9)

Table S2. Demographic and clinical characteristics of the replication cohort. The values are the mean with the standard deviation shown in parentheses (SD) or the absolute number with the percentage of patients shown in parentheses. The value for interstitial fibrosis is the percentage of the biopsy sample showing fibrosis, with 0% being no fibrosis and 100% being complete fibrosis. The value for lymphocytic infiltrate was scored by histology analysis of the biopsy samples (0-3). Abbreviations: eGFR, estimated glomerular filtration rate; BMI, body mass index.



Name	Sequence (5’-3’)
TNF DMR sgRNA1 F	CACCGGCTCATGGTGTCCTTTCCAG
TNF DMR sgRNA1 R	AAACCTGGAAAGGACACCATGAGCC
TNF DMR sgRNA2 F	CACCGCAGCAGGCAGAAGAGCGTGG
TNF DMR sgRNA2 R	AAACCCACGCTCTTCTGCCTGCTGC
TNF DMR sgRNA3 F	CACCGGGGATGGAGAGAAAAAAACG
TNF DMR sgRNA3 R	AAACCGTTTTTTTCTCTCCATCCCC
TNF NC sgRNA1 F	CACCGAAGTAGGCATGAGGGATCAC
TNF NC sgRNA1 R	AAACGTGATCCCTCATGCCTACTTC

Table S3. Guide RNA primer sequences.



Name	Sequence (5'-3')
Human *BAX* F	ACCAGCTCTGAGCAGATCATGA
Human *BAX* R	TGCAGCTCCATGTTACTGTCCA
Human *BAK1* F	GGTTTTCCGCAGCTACGTTT
Human *BAK1* R	TGGTCTGGAACTCTGAGTCATAGC
Human *BID* F	CCTCCAAAGCTGTTCTGACAAC
Human *BID* R	CCTGGCAATATTCCGGATGA
Human *TP53* F	TCATCTTCTGTCCCTTCCCAGA
Human *TP53* R	GTGTGGAATCAACCCACAGCT
Human *RIPK1* F	CACAGAGAAGTCGGATGTGTACAG
Human *RIPK1* R	TTCCCAGCAGAGCTTCATGA
Human *TNF* F	TGCTTGTTCCTCAGCCTCTTCT
Human *TNF* R	GGTTTGCTACAACATGGGCTAC
Mouse *Bax* F	CCACCAGCTCTGAACAGATCAT
Mouse *Bax* R	CAGCTCCATATTGCTGTCCAGT
Mouse *Bcl2* F	CCTGTGGATGACTGAGTACCTGA
Mouse *Bcl2* R	GTATGCACCCAGAGTGATGCA

Table S4. qPCR primer sequences.

References

1. Safran M, Chalifa-Caspi V, Shmueli O, Olender T, Lapidot M, Rosen N, Shmoish M, Peter Y, Glusman G, Feldmesser E, et al. Human Gene-Centric Databases at the Weizmann Institute of Science: GeneCards, UDB, CroW 21 and HORDE. *Nucleic acids research.* 2003;31(1):142-6.
2. Kang HM, Ahn SH, Choi P, Ko YA, Han SH, Chinga F, Park AS, Tao J, Sharma K, Pullman J, et al. Defective fatty acid oxidation in renal tubular epithelial cells has a key role in kidney fibrosis development. *Nature medicine.* 2015;21(1):37-46.