Diabetic Complications Consortium

Application Title: Integrative 'Omics' Analysis of Progression of Renal Decline in Type 1 Diabetes

Principal Investigator: Marcus Pezzolesi (PI), Zhou Chen (Co-I)

1. Project Accomplishments:

As part of this project, we establish DNA methylation (DNAme) on 68 individuals with Type 1 diabetes from the Joslin Kidney Study (JKS), including 34 patients who developed end-stage kidney disease (ESKD) and 34 patients who did not. These data we jointly analyzed with data from these same 68 individuals, collected as prior timepoints, to establish persistent DNAme signatures at ESKD-associated cytosines 5′ to guanines (CpGs) sites.

2. Specific Aims:

Specific Aim 1. Establish DNAme profiles and gene expression profiles in well-phenotyped participants of the JKS.

We profiled DNAme on archived whole blood (WB) DNA samples collected during follow-up examinations (1-14 years following baseline) from 68 JKS participants who had prior DNAme profiles established on baseline samples. These 68 samples are a subset of 277 participants of the JKS (52% of whom progressed to ESKD) who had baseline DNAme profiles previously established. Both baseline and follow-up DNAme variation were profiled using Illumina's Infinium MethylationEPIC BeadChip (EPIC) arrays. All selected JKS participants had T1D (mean HbA1c=8.94%) and documented diabetic kidney disease (DKD) with proteinuria and/or impaired renal function (median ACR=728.90 mg/g and mean eGFR=62.90 mL/min/1.73m^2). None had ESKD at baseline. During the follow-up period (7-20 years), the participants experienced different rates of kidney function decline, represented by eGFR slope, with half (n=34) progressing to ESKD. DNAme at more than 840,000 CpGs were reliably profiled following quality control of these data.

Due to limited availability of whole-blood RNA specimens from the selected cohort (only ~20% had available specimens), we chose to expand the original proposed sample size of our DNAme studies by 25% to include additional JKS participants.

High-quality genome-wide genotyping data, generated as part of the JDRF Diabetic Nephropathy Collaborative Research Initiative, was available for 264/277 of the JKS participants of this study. SNPs with a minor allele frequency > 0.05 and with imputation information scores > 0.30 were included in the analyses. For ESKD-associated CpG identified in this cohort, association analyses were performed using SNPtest on an additive model adjusting for gender, baseline

variables (age, duration of diabetes, log-eGFR, log-ACR, log-HbA1c, hypertension), cell compositions, and batch to identify methylation quantitative trait loci (meQTLs).

Specific Aim 2. Identify persistent DNAme signatures, methylation quantitative trait loci (mQTLs), and expression QTLs (eQTLs) associated with rapid renal decline.

DNAme data from 277 JKS participants were used to identify CpGs associated with the risk of EKSD development. Among the 846,816 CpGs whose DNAme levels were reliably measured across all participants, we first identified 165 CpGs associated with ESKD using the Cox Proportional-Hazard (coxPH) model without any adjustment at $P < 5e\text{-}5$. Next, to identify CpGs whose DNAme levels improve ESKD risk prediction, for each of the 165 CpGs, we next performed coxPH analysis after adjusting for major baseline demographic/clinical predictors selected by LASSO regression (i.e., eGFR, ACR, HbA1c, age, hypertension status, and stratified by sub-cohorts). At false discovery rate (FDR) < 0.05, a total of 17 CpGs were identified to be associated with risk of ESKD development independent of the adjusting variables, designated as ESKD-associated CpGs (**Figure 1**).

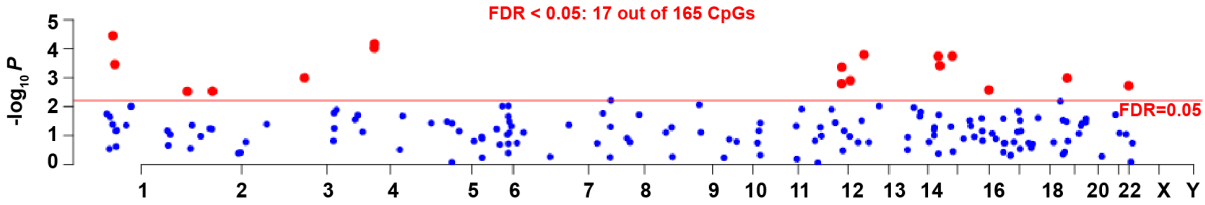

Fig. 1. Identification of CpGs whose DNAme level is associated with ESKD-risk. Manhattan plot showing the association of DNAme at the CpGs associated with ESKD risk ($P < 5e\text{-}5$) after adjusting for baseline eGFR, ACR, HbA1c, age and hypertension, and stratified by sub-cohorts. CpGs that remained significantly associated with ESKD risk at FDR < 0.05 are termed ESKD-associated CpGs.

DNAme at 5 ESKD-associated CpGs showed negative association with ESKD-risk, i.e., high methylation at these sites was associated with lower risk of ESKD [Hazard ratio (HR) < 1.0]. These include cg07161341, cg02250900, cg20978937, cg15261665 and cg18008766 in *TBC1D1, DMXL2*, *PLD4*, *LTF*, and *SRSF7* respectively. DNAme at the remaining 12 CpGs showed positive association with ESKD-risk, i.e. high methylation at these sites was associated with higher risk of ESKD [Hazard ratio (HR) >1.0]. These CpGs included: cg12075771, cg25564144, cg00646813, cg07663788, cg10212605, cg10779340, cg23029183, cg23069305, cg10667857, cg07364018, cg17944885, and cg07969217 in/near *EPRS1*, *RBFOX2*, *TLR1*, *SSH1*, *WDR25*, *ZBTB40*, *CACNB3*, *PFKM, LONP2, RNU1-2*, *ZNF788P*, and *TPH2* respectively.

As part of this project, WB DNAme profiles by EPIC arrays were obtained on 68 out of the 277 JKS participants from above using a second set of WB DNA samples collected at a different time-point of 1~14 years of gap (median: 4.7 years; 25th-75th percentile: 3.1~8.9 years). DNAme

datasets including both samples of the 68 participants were preprocessed using the same approach. Genome-wide comparison in P2 (samples collected at later point) vs. P1 (collected at earlier time point) was performed using paired t-test on DNAme datasets after adjusting for batch effect. Remarkably, when we examined the DNAme profiles from the second set of WB samples collected at a different time point from 68 out of 277 participants, we found DNAme at all the 17 ESKD-associated CpGs remained stable without change over 1-14 years, except for just a couple of CpGs depicting very mild change at nominal $P < 0.05$) (**Figure 2**). Hence, our identified ESKD-associated CpGs could be considered stable biomarkers for ESKD-risk prediction.

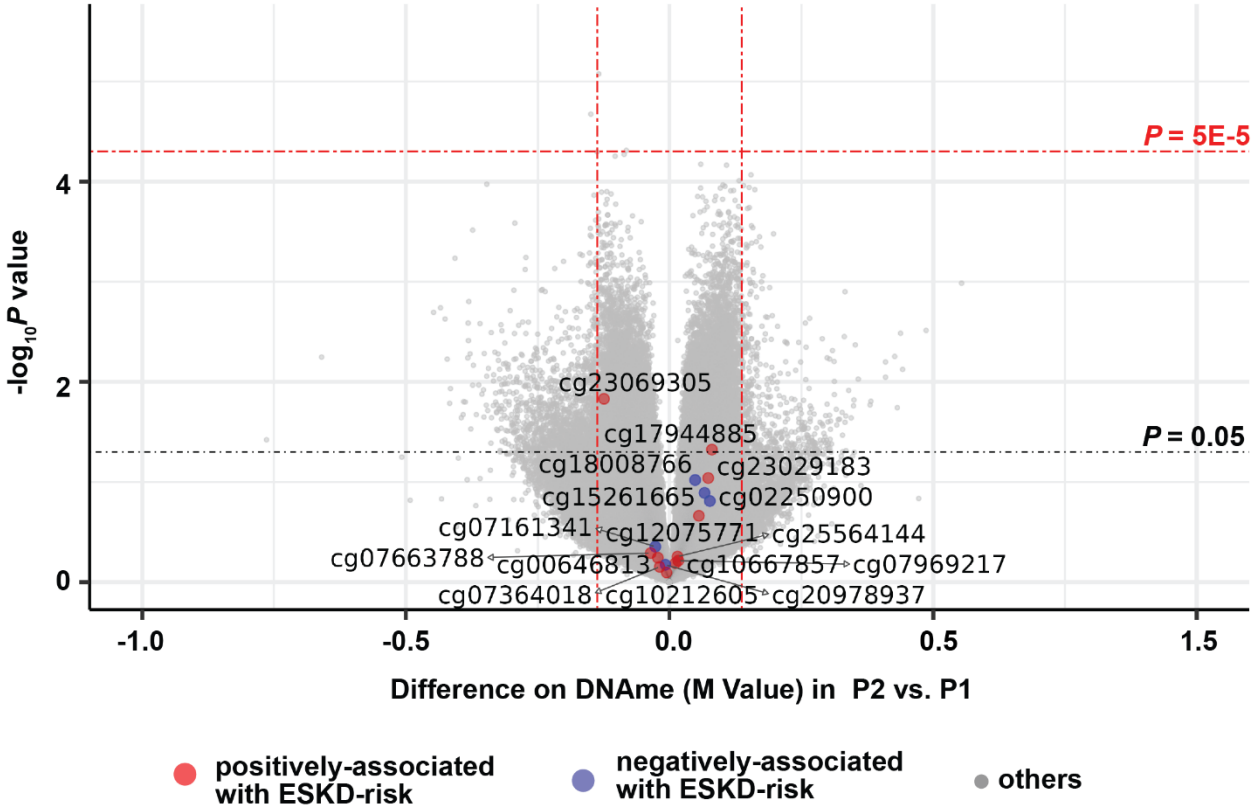

Figure. 2. The stability of DNAme at ESKD-associated CpGs. WB DNAme was profiled on 68 participants of the Joslin Kidney Study (including 34 patients who developed ESKD and 34 patients who did not) using another set of WB DNA samples from same participants and by same platform (EPIC arrays) with a gap of 1~14 years. Genome-wide comparison in P2 (sample collected at later time point) vs. P1 (sample collected at earlier time point) across the 68 patients was performed using paired t-tests on DNAme datasets after adjusting for batch effect. The results are presented as volcano plot with each dot representing one CpG. The ESKD-associated CpGs (total 16 reliably covered in the dataset) are highlighted as red dots for ESKD-risk positively-associated CpGs and blue dots for negative-associated CpGs.

Next, to map genetic regulators of DNAme that are associated with ESKD, we next performed meQTL analyses on the ESKD-associated CpGs using genotypic data available from 264/277

JKS individuals. We identified 7 CpGs having at least 5 meQTL single nucleotide polymorphisms (SNPs) at *P*<5e-8. Regional association plots covering *cis*-meQTL SNPs for cg18008766, cg00646813, cg10212605, cg07663788, cg20978937 and cg07364018 (**Figure 3A-3F**), or *trans*-meQTLs for cg17944885 (**Figure 3G**) demonstrated a cluster of SNPs with moderate/high linkage disequilibrium near the lead SNP at each locus (purple diamonds, Figure 3), confirming true association signals detected in our analysis.

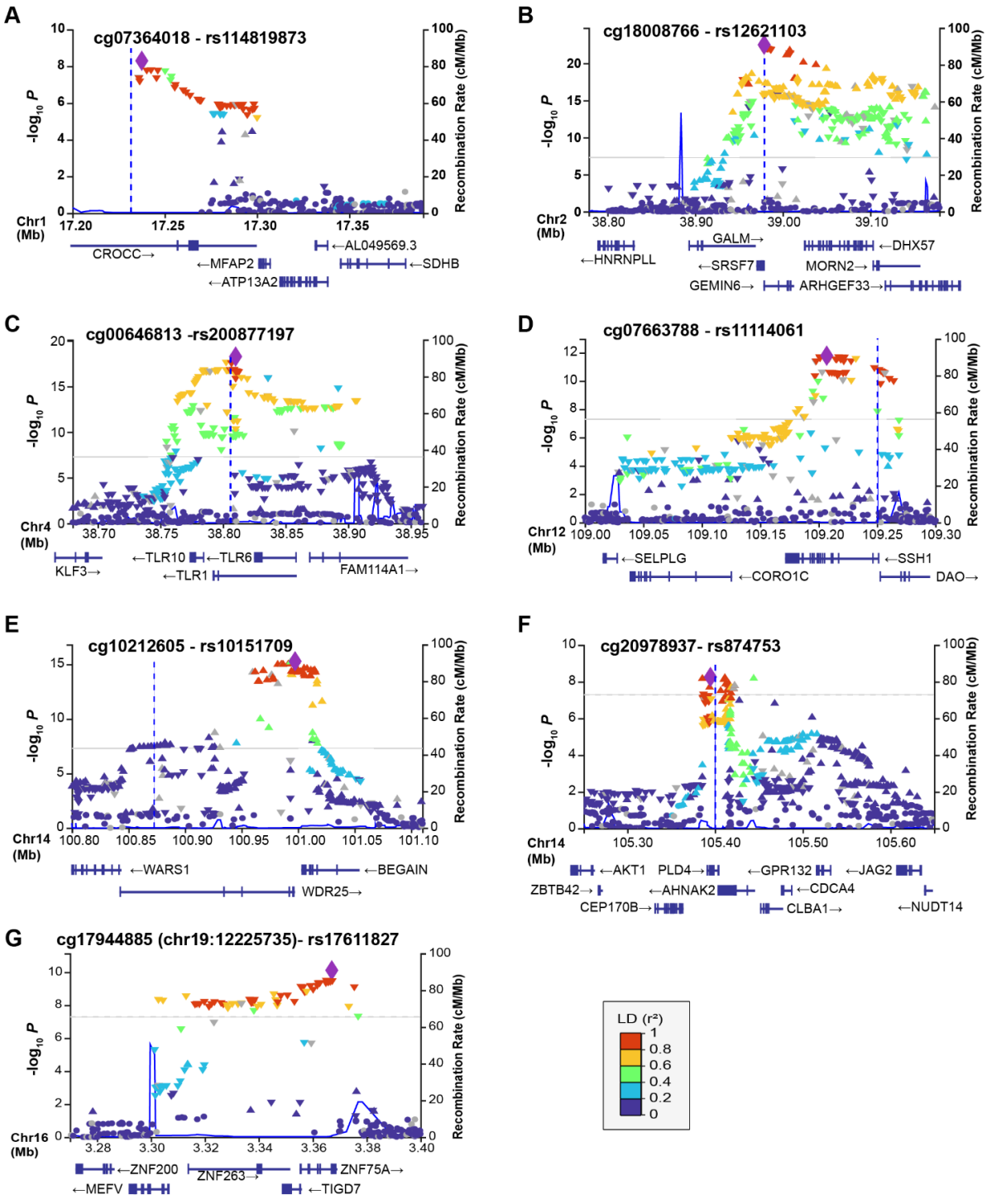

Fig. 3. The association of DNAme at ESKD-associated CpGs with genetic variations. For each ESKD-associated CpG, meQTL analysis was performed using DNAme and available genetic data from JKS participants (n=264/277). **(A-F)** Regional association plots show association between DNAme at one CpG whose ID is shown on top of the plot (location presented by dashed line) and the nearby SNP genotypes. **(G)** The association of DNAme at cg17944885 with its trans-meQTL SNPs. In each plot, each dot represents one SNP. Y- axis: The left axis shows the association *P* value between SNPs and DNAme in $-\log_{10}P$ format. The right-axis shows the recombination rate. X-axis shows chromosome coordinates per hg19. Purple diamond: The most significantly associated SNP (labeled with rsID) for each CpG. Color scale: Linkage disequilibrium (LD, measured based on the 1000 Genomes November 2014 European population) between the significantly-associated SNP and other SNPs.

Finally, to determine links between these genetic variants and DNAme, we applied *in-silico* analysis and identified 10 transcription factors (TFs) having at least 5 binding sites impacted by the identified cis-meQTL SNPs. These TFs are known to be modulated DNAme (ZNF263, IRF1 and EGR2) and/or involved with kidney disease, including inflammation (IRF1, ARID3A, SP1, MZF1, KLF5, and EGR2) and kidney fibrosis (IRF1, SP1, SP2, MZF1, PAX5, KLF5, and EGR2). Among these, the most abundant binding sites were for ZNF263 impacted by *cis*-meQTL SNPs of 4/6 CpGs. Interestingly, *ZNF263* gene is located at the *trans*-meQTL region of cg17944885, suggesting it may also be involved in trans-effects of genetic variants on DNAme. These results indicate key genetic variations may drive DNAme changes at some ESKD-associated CpGs by affecting the binding of TFs involved in DNAme and DKD, suggesting both direct and indirect effects of these genetic variants on ESKD development.

3. Publications:

1. Whole Blood DNA Methylation Signature, Circulating Proteins and Risk of Progression to End-stage Kidney Disease in Type 1 Diabetes, Zhuo Chen, Eiichiro Satake, Marcus G Pezzolesi, Devorah Stucki, Anna Syreeni, Zaipul I Md Dom, Adam T. Johnson, Jaxon B. King, Xiwei Wu, Emma H Dahlström, Per-Henrik Groop, Niina Sandholm, Andrzej S Krolewski, Rama Natarajan, Science Translational Medicine, *In Review*